THE UNIVERSITY
of EDINBURGH

What to Talk About, and How:

Studies on Prominence and Patterns of Coreference

Luca Bevacqua

Thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

2020

Declaration

I declare that this thesis has been composed by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where stated otherwise, the work presented is my own. The contributions to co-authored papers are explicitly acknowledged in the introduction of each chapter in which they are included. The reproduction of the papers is compatible with the open-access publication policies of the original collocations; permission has been obtained from the co-authors.

Luca Bevacqua

December 2020

Abstract

The concept of prominence has been variously defined, and it overlaps with other ideas in both theoretical and cognitive linguistics, such as activation, emphasis, or accessibility. Moreover, prominence has an important role in the interpretation and production of language, influencing what anaphoric patterns are produced and/or seen as mostly likely, and what referring expressions are chosen to express coreference.

This thesis presents psycholinguistic, crosslinguistic studies on prominence and coreference, grouping them in two parts respectively on the surface form and repercussions of prominence and on prominence as seen in different components of meaning. The first study, on English, surveys how prominence is expressed in cleft constructions by extracting emphasis markers and "formal" features within clefts from two corpora at different registers, exploring the patterns in which syntactic marking, graphical emphasis markers, and the variants of contraction, pronoun and complementiser are used in a synergy to express prominence. The second study uses the same structure of the cleft in Italian, and focusses on two factors affecting prominence: information structure and sentence boundary. It then analyses the next-mention choices that writers make, and how this choice is carried on with referring expressions.

Moving to prominence in smaller linguistic components, the studies in the third section analyse event and entity coreference in English, French, German, Italian, and Spanish, using different referring expressions and features of the verb (aspect and causative-inchoative alternation) as proxies to manipulate the prominence of entities versus the events in which they are involved. Finally, the fourth and last section investigates number conceptualisation in named entities in the same five languages: in coreference, speakers have to choose whether to index the entity according to its morphosyntactic or notional number, marking agreement on the pronoun consequently. The prominence of grammatical and semantic number in the speakers' indexing of referents is shown to change crosslinguistically and with the formality of a text, as well as with features of the entity.

Overall, the results of this research show a varied interplay between prominence and patterns of coreference, with different manifestations at different levels of linguistic structure and results that can sometimes be extended crosslinguistically.

Lay Summary

When communicating, speakers need to continuously make two choices: what to talk about, and how to talk about. Not only one can change the topic while speaking, but these topics can be named in many ways: for example, the phrases "James Joyce", "the writer" or "he" can all refer to the same person. Such choices are influenced by many factors, and this thesis focusses on one of them: prominence. Prominence is the property of an entity that is at the centre of attention, and can be defined in both cognitive and linguistic terms.

This thesis presents work across five languages (English, Italian, French, German, and Spanish) and uses an experimental method to explore the relation of prominence with what speakers choose to say and how they choose to say it. The studies presented are divided in two groups: first, prominence is examined in how it appears on the surface and on the linguistic repercussions it has, second, prominence is studied in different components of meaning.

In the first part, two studies are presented. The first study surveys how prominence is expressed in English in Twitter, a very informal genre of language: different features are examined in the context of a construction called cleft (e.g.:"It is Mark who called Mary on the phone"), such as graphical elements used to express emphasis (e.g. *askerisks*), and the use of these features in Twitter is compared to that in a more standard variety of English. The second study focusses on the same construction but in Italian, and explores how this structure, which is used to mark prominence, along with other factors influence the speakers' choices of who to talk about and how.

The second part is also composed of two studies, on the five languages. The first study looks at how speaking about an entity (e.g. "the flower") is different than speaking about an event (e.g. "the flower bloomed"), and at what influences the choice between talking about one or the other. The final study is about how named entities, entities that represent a multiplicity of parts, are talked about: in the plural (e.g. "Pink Floyd were a rock band"), or in the singular ("Pink Floyd was a rock band"), looking at whether their parts or the sum of them is more prominent.

Overall, the research shows a rich interplay between prominence and the choices speakers make about what to talk about and how to talk about it, and this interplay can be seen at various level, from the surface to the way something is perceived in its parts.

Acknowledgements

First, I want to thank my supervisors: Hannah Rohde and Chris Cummins. Thank you for all the feedback and comments and prompt answers to my many questions. Thank you Hannah for creating the opportunity for me to work with you and other people – a collaboration whose results form two chapters of this thesis. I would also like to thank Patrick Sturt and Roger van Gompel for accepting to examine my thesis.

Next, I want to thank all the other people I collaborated with: Christian Hardmeier & Sharid Loáiciga for the brainstorming sessions, experiments (and papers), and following them up with more experiments (and papers) on two research trajectories that were not in my plans at first but proved fruitful and, what's most important, interesting. I want to thank the people in project A03 of the SFB 1287 at Univerität Potsdam for hosting me for a few months and taking interest in my work: Tatjana Scheffler, Manfred Stede, Berfin Aktaş & Yulia Clausen. Thank you Tatjana for convincing me to turn the research into a paper, and for writing it with me.

To all my friends across the years & the continents & everyone else who helped in small or big ways: thank u! Thank u to all writers, musicians, artists & poets.

Finally, to my family for always supporting me: grazie.

Contents

II. The Prominence of Components of Meaning

Chapter 1 Introduction

Referring to things, be they people, objects, events or concepts, is one of the most important functions of language, and, as such, it comes with its burden of choices to make.

First, a speaker needs to decide what they want to talk about, presumably following a coherent line of thought according to which whatever was being referred to is relevant in a linguistic or extralinguistic context. This choice is central to sharing the object of a speaker's communication to a listener.

When this first choice is made, they will need to find a way to express it. This means choosing the words that will refer to the matter at hand thanks to some linguistic convention: not only choosing what to talk about, but also how to talk about it. We can assume that in most cases the second of these choices follows the first.

Both choices are related to the concept of prominence, in the specific sense of how prominent the talked-about element is both linguistically and conceptually. This thesis deals with this concept in two ways: first, it looks at how prominence is manifest at the surface-level of a text, considering what tools are used to mark it, and at what consequences this marking has on the choice of what to talk about and how to talk about it, that is the selection of antecedent and referring expression. Second, the thesis investigates prominence in different components of meaning: the prominence of whole events versus entities involved in an event, and the prominence of the grammatical and notional (conceptual) number of named entities. The studies take a psycholinguistic and multilingual approach to compare the predictions of mainly English-centric theoretical frameworks, working empirically on English, Italian, Spanish, French, and German.

Referring to something – the referent – with a certain expression means identifying something in the world and putting it in relation to a linguistic expression. The choice of referring expression follows some shared principles and interlocutors do not need to constantly renegotiate it. This contract between speakers can be a stricter or looser one: while some words mean something to various degrees of univocity (like "cat", or even more stringently "Madrid"), others are less strict and more context-dependent: this is especially true

of pronouns, which can even not audibly express any meaning other than the linking function when manifested as phonetically void particles in languages such as Italian or Korean. All these expressions serve the purpose of referring to something and/or linking back (or forward) to another referring expression within the text: this linking function is called anaphora (or cataphora for forward linking), and the expression fulfilling the task is termed anaphor (or cataphor). The main role of these links between expressions, and the interpretation of expressions as different ways to refer to the same thing, is to participate in creating cohesion within a text (Halliday and Hasan, 1976; Hobbs, 1979).

Both the choice of who to talk about and that of how to talk about it are influenced by how prominent the referent is. The prominence of a referent has been defined in various ways and called by various names: most commonly, salience (von Heusinger, 1997), accessibility (Ariel, 1990), activation (Lambrecht, 1994; and in more cognitively oriented frameworks such as Gernsbacher, 1990; Almor, 1999), or using terms from the vocabulary of information structure like focus and topic. Some of these definitions are directly used to explain, or at least related to, information structure: for example, Lambrecht (1994) uses the term activation in relation to information structure and its effect on the mental representation of discourse, while Zimmerman (2008) relates emphasis to contrastive focus. These definitions also focus on different manifestations of prominence, whether more prosodically, semantically or syntactically. Chapter 1 of this thesis looks at how prominence manifests in English in Twitter text – a versatile form of writing that employs both prosody-like markings and syntactic tools to mark emphasis.

In some cases, it is intuitive that a referent that is more prominent because the discourse revolved around it will continue to be talked about, rather than the speaker shifting and choosing something else to talk about. Consider the following (constructed) paragraph:

1) James Joyce wrote most of the short stories included in "Dubliners" around 1904. He then wrote the closing novella of the collection in 1907. He had to fight censorship and delay the publishing of the stories multiple times, but he finally published the book in 1914.

The paragraph is quite straightforward in being about Joyce himself and a list of actions he has done: first he wrote most of the short stories in "Dubliners", then he wrote The Dead (referred to via the noun phrase "the closing novella"), after which he had to fight publishers over his use of real fact and places and allegedly obscene language, and finally he managed to publish the book. After each sentence, it is easy to just keep mentioning him since he is the protagonist of the text and simply refer to him with a "he" after using a full name at the very start; in the same way, Dubliners shifts its form to others, that is "the collection", "the stories" and "the book". These subsequent expressions are more ambiguous than a proper name, but since their referent has been made clear in the preceding portion of the text they can be used with a low chance of misunderstanding. Moreover, the use of a noun phrase such as "the stories" can also add some new information that was not already included in the antecedent's phrase.

Note that, as shown in (1), coreference can happen within or between sentences. The function is the same but, the more distant the two expressions are, the more changes occur in the common ground in between their occurrence. The common ground can be defined as the set of assumptions that the speaker (or signer or writer) and the receiver of the communication share, including the world knowledge assumed to be shared, the context of the communication, and the topic of the discourse (Allan, 2013). The common ground is stored or at least readily available in short term memory and processed and updated over time, with new information coming to the table and new commitments being made by both parties to the communication (e.g. Brown-Smith and Duff, 2016; Geurts, 2019).

Prominence thus shifts sides in the course of a discourse depending on what is at issue in each sentence related to its preceding sentence. A common assumption in the literature is that prominence and its changes have the same effect on both the choice of which referent to talk about and the choice of the referring expressions used to re-mention it. This will be investigated more in depth in Chapter 4 of this thesis: this presents a language production experiment in Italian testing how information structure and sentence boundary (to features influencing prominence) affect antecedent re-mention and choice of referring expression. Specifically, the study uses cleft sentences to mark a referent as new or

contrastive information. The results show how focusing a referent, making it more prominent but at the same time marking it as new information, has apparently opposite effects on the choices of next mention and referring expressions: focussed referents are selected less for re-mention, but referred to with light expressions, that is, expressions that are either semantically or phonologically weak (or both, such as null pronouns).

Sometimes the choice of referent is not as obvious as in paragraph (1). Consider the incipit of *The Dead*:

2) Lily, the caretaker's daughter, was literally run off her feet. Hardly had she brought one gentleman into the little pantry behind the office on the ground floor and helped him off with his overcoat than the wheezy hall-door bell clanged again and she had to scamper along the bare hallway to let in another guest. It was well for her she had not to attend to the ladies also. But Miss Kate and Miss Julia had thought of that and had converted the bathroom upstairs into a ladies' dressing-room. Miss Kate and Miss Julia were there, gossiping and laughing and fussing, walking after each other to the head of the stairs, peering down over the banisters and calling down to Lily to ask her who had come.

It is hard to predict who will be mentioned next for a series of reasons. On the one hand, Lily seems to be at the centre of the scene for the first half of the paragraph. She's introduced with her name and a noun phrase describing who she is ("the caretaker's daughter"), and she is kept at the centre of attention by talking about her actions. On the other hand, other characters are introduced: Miss Kate and Miss Julia, for example, are also introduced by their name, and the latter two sentences of the paragraph revolve around their whereabouts. They get instantiated and activated in the mind of the reader by being talked about. Towards the end Lily is mentioned again, bringing her from short term memory to an activated status.

Activation is, in a psycholinguistic sense, the feature that characterises a referent that is not only identifiable (that is, known by the recipient of the communication), but also at the centre of attention, brought to the forefront of consciousness (Chafe, 1987; Lambrecht,

1994). On the middle ground between an inactive and an active status are accessible representations, which are peripheral in one's consciousness. These concepts are linked to memory: inactive representations are known but stored in long-term memory, accessible representations in short-term memory, and activated representations are the current centre of attention. Any time a referent is mentioned it is brought back to an active state, and then its status starts to decay over time and move towards the periphery and out of consciousness. The way in which a referent is mentioned is relevant both to its successful activation and to whether it is made prominent. This is, again, a fundamental interest of Chapter 3, which uses Twitter data to investigate multiple ways in which a referent can be marked as prominent: syntactically in clefts, via emphasis markers, and with other "formal" features of the sentence such as contraction in the cleft, pronoun case and choice of complementiser. The findings of the study show how these prominence-signalling tools are often used concurrently to mimic intonation and mark a referent with a strong emphasis. Moreover, the study finds that the case of the clefted pronoun does not correlate with the syntactic role of its referent.

Looking back at paragraph (2), it is fair to say that starting a new sentence with either of the above-mentioned human referents would feel quite natural. A fourth human referent evoked in the scene (the caretaker) could also be picked up, despite his very peripheral position in the discourse. Compare the two sentences in (3):

3) a. The caretaker himself was sitting in the great hooded chair in the hall, observing the party unfold.
   b. He was sitting in the great hooded chair in the hall, observing the party unfold.

While sentence (3a) would make the shift in reference clear using a strong definite description to signal the change ("the caretaker himself"), sentence (3b) is much more ambiguous, leaving reference unspecified. In a normal conversation, an underspecified sentence like this would risk ambiguity, unless context helped identify the subject of the utterance: context, being part of the common ground, can effectively supplement a text's information. The task of identifying which referent an expression refers to, linking it to a

previous (or following) element in the text, is broadly called coreference resolution in both linguistics and natural language processing, and it is a well-studied phenomenon in psycholinguistics – especially for some specific cases in English (for a review: Almor and Nair, 2007; Rohde, 2019; see also Poesio et al., 2016, for a review of the computational state of the art).

In example (3b), the caretaker was simply not prominent (or activated, or salient, etc.) enough to be brought up with a minimal referring expression such as a pronoun: a more informative referring expression is needed to unequivocally move its representation from the periphery of consciousness to the centre of attention. Talking about the caretaker would have demanded a switch in referent to change the topic of the discourse from the more central characters to himself. The inventory of possible referring expressions includes any noun phrase, such as proper names, nouns, but also pronouns and gaps, and they can be ordered in a scale of decreasing informativeness (e.g. Ariel, 1990): while a full noun phrase points to a referent through its meaning, pronouns only give some minor information such as number and gender, or even none at all in the case of a null form.[1]

Suddenly talking about a character who had been peripheral in memory (or even not known at all) up to that point is not the only thing a speaker could do. The speaker could talk about a non-human referent, like a piece of furniture (4):

4) The stairs were thickly carpeted and accompanied each step with a soft thud.

All the previous examples are in the third person, using pronouns "she/he/it". Human third-person singular referents are the most studied psycholinguistically, as a prototypical case of coreference resolution. More abstract referents, such as an event or the setting of the scene, are less studied, as there are more confounding factors that need to be controlled, such as the structure of an event or the number or entities involved in it. Reference

[1] Depending on the language, some information can be encoded in other parts of speech, such as verbs or adjectives.

to an event is exemplified in (5), at the beginning of the next paragraph in the Joyce novella, with the pronoun "it" (used as a cataphor in the first sentence and as an anaphor in the two subsequent ones) referring to the "annual dance":

5) It was always a great affair, the Misses Morkan's annual dance [...] Never once had it fallen flat. For years and years it had gone off in splendid style, as long as anyone could remember.

The use of anaphors to refer to non-human referents is less studied (but see e.g. Betancort et al., 2009, on the influence of animacy on syntactic processing and Fukumura and van Gompel, 2011, for its effect of referring expressions), and most experimental research has adhered to designs in which two stereotypical human referents are used, to have as few confounding factors as possible. Chapter 3 of this thesis reports two multilingual story continuation experiments manipulating referring expression and verb semantics to evaluate their effects on entity vs event coreference. Previous evidence suggests that events, being more semantically complex and harder to introduce directly, will tend to be referred to with demonstratives rather than personal pronouns in English (Hedberg et al., 2007). Demonstratives additionally have often been characterised as rejecting the most prominent antecedent and signalling prominence shift (see e.g. Givón, 1983; Comrie, 1997). These ideas are confirmed crosslinguistically by the studies' results, showing that semantically richer and phonetically heavier referring expressions are interpreted to corefer to events, which are more complex antecedents. Moreover, the complexity of an event's structure also influences reference: adding implicit entities via the use of a verb in the causative-inchoative alternation (e.g. "the snow melted", implying that something like warmth melted the snow) inhibits the prominence of the explicit referent, thus also yielding more event coreference.

Non-entity antecedents include not only events, which are themselves composed both of entities and of relations between them (expressed by a verb phrase), but also textual elements such as full clauses or rhetorical moves. This case is not considered in the present work, but, for example, after saying paragraph (2) out loud one could continue talking about it as a piece of writing metatextually (6a) or metalinguistically (6b):

6) a. This is the incipit of my favourite short story, *The Dead* by James Joyce.
   b. This is what is called a paragraph: it's a section of writing separated by a new line and indentation, and it's used to deal with a single theme.

The availability of such a rich and complex array of possible antecedents overviewed thus far makes the challenge of fully describing coreference resolution even greater, and the fact that humans effortlessly resolve reference baffling.

Finally, even entities can be seen as a collection of smaller elements, rather than a unitary referent. A relatively simple entity like "Dubliners" can be understood as a single "book" object, a collection of stories, or as a sum of otherwise defined individual parts (genres, themes, paragraphs, words...). This can be seen more obviously when referring to named entities such as names of companies. These names can be invented or repurposed words (e.g. Google, Apple) or acronyms (IBM), and can agree with the verb and subsequent anaphors via their grammatical number (e.g. *Ikea is a Swedish company. It sells furniture*) or via their notional number, being perceived as a group of individuals (*They sell furniture*). This preference may not only be crosslingual or dialectal, but also be dependent on other, more nuanced variables, both language-internal, such as formality and genre, and lexical-semantic, such as the features of an entity (Levin, 2001). Chapter 6 of this thesis reports two crosslinguistic studies on named entities, testing what component of named entities of different type (like companies, music bands and sport teams) is more prominent: their grammatical number which characterises them as a single complex entity, or their notional number, cognitively picturing them as a group of many elements. The findings show that different languages have different biases when co-referring to named entities, with a further effect of formality (at least in English), and that different types of entities foreground either the unity of the entity or their composition from many individual parts (e.g. respectively companies vs music bands).

Reference can thus be very complex, of different kinds, and the main referent that we are talking about can change swiftly and often without jeopardising the coherence of the text. However, in the vast majority of cases reference is resolved naturally in a conversation,

with few instances of misunderstandings and corrections needed. Such an array of possibilities and links between referents or referring expressions has been studied in linguistics under many frameworks, from Chomsky's syntactic theories to semantics and pragmatics perspectives, with theoretical, computational, and experimental methods (for a review of theoretical perspectives, see Huang, 2000; Botley and McEnery, 2000, for a more method-oriented overview). Most theories try to grapple with a specific challenge, often working in a specific language and at times extending the results on other languages assumed to follow the same patterns (an example is Carminati's, 2002, theory being often extended to pro-drop languages in general). From a theoretical point of view, it is hard to incorporate knowledge and evidence from the various fields of linguistics into a single theory of anaphora, and from an empirical perspective there are limits to the number of cases and variables that can be investigated in a study, forcing experiments to stick to some sanitised cases with as few confounding factors as possible. Finally, corpus and computational linguistics have struggled not only with non-uniform annotation standards, but also with the scarceness of non-English data.

This thesis takes an experimental and crosslinguistic point of view. A set of classic theories of prominence and coreference resolutions are considered and assessed with regards to the findings from psycholinguistics experiments in five European languages: English, French, Italian, German, and Spanish. This approach aims at testing theoretical predictions such as Ariel's (1990, 2001) Accessibility Theory and Carminati's (2002) Position of Antecedent Strategy among others, on their target languages, namely English and Italian, but also at checking whether those predictions accurately describe patterns observed in other languages within the same language families: for example, results on Italian have often been generalised to cover (mainly peninsular) Spanish, and vice versa. Finally, starting from the same theoretical considerations to empirically study different languages offers the possibility to think about the very definitions at the base of theories, and especially the characterisation of "prominence" as something cognitive rather than language-specific.

In the next chapters, first I present the theoretical perspectives and review prior psycholinguistic evidence. Psycholinguistic studies have both targeted specific frameworks to

support them, and specific phenomena of coreference (most notably, coreference resolution). Most of this work predictably focusses on English, and sometimes the predictions are assumed automatically to translate to human language in general, without proper experimental testing.

Then the studies making up this thesis are divided in two parts: work on the surface form of prominence and its repercussions on the election of an antecedent for re-mention and the choice of referring expression; and prominence in different components of meaning, namely that of entities or whole events as antecedents and that of the grammatical versus notional number of named entities.

Chapter 2 Review of the literature

Coreference and pronoun resolution, as well as the use of referring expressions, are among the most studied phenomena in linguistics, both theoretically and empirically. The concept of prominence, variously named and defined, is often employed in an attempt to motivate the coreferential patterns systematically, computationally or cognitively. In this section, some of the most influential theories dealing with these topics are reviewed at a general level, to preface the literature reviews specific to the studies presented in each of the chapters.

First (§2.1), two highly influential frameworks which form the base of many theories are introduced: Chomsky's Government and Binding theory (1981), and the Gricean programme (1989). These theories respectively provide the syntactic and pragmatic foundations to the study of coreference and referring expressions. Next, subsequent models are presented, such as some neo-Gricean perspectives (§2.2), models that associate the reduction in referring form with the predictability of the antecedent, most notably Accessibility Theory (§2.3), cognitive models with explanations based on salience and accessibility, memory and coherence (§2,4), and finally newer frameworks moving away from some common assumptions (§2.5).

2.1 Fundaments: Chomsky and Grice

The principles which allow reference to happen and anaphors to link to other referents in a text have been studied under different frameworks. *Anaphor* is here used in a broad sense, unlike in the strictly structural sense of Government and Binding theory (Chomsky, 1981): a word or phrase whose interpretation depends on context and cotext, interpreted through a link to another word or phrase. The word is used as an umbrella term for both proper anaphor, referring back to a previous element (e.g. in the sentence *Rimbaud*$_1$ *wrote A Season in Hell when he*$_1$ *was nineteen*), and cataphor, referring to a subsequent element (e.g. *When he*$_1$ *wrote the Illuminations, Rimbaud*$_1$ *was twenty*).

Two very different perspectives are foundational: Chomsky's Government and Binding theory (1981) from a structural point of view, and Grice's conversational principles (1989) for a discourse pragmatics point of view. While the first contributes to a generative, syntactic point of view in which anaphoric relations are explained by the derivation of a sentence, the second offers a discourse-centred standpoint, where the relations are a product of the speakers' cooperation in a conversation.

Chomsky's (1981) binding theory at its core employs syntactic and semantic principles to restrict the distribution of noun phrases (NPs) that can anaphorically refer to the same referent. Given that it is a theory of language in an abstract, mentalist sense (Universal Grammar), its rules can be applied to any human language.

The distribution of referring expressions is described by distinguishing NPs into three categories explained by two essential morphosyntactic binaries: [±anaphor] and [±pronominal]. Referential expressions, which in the framework include common nouns and proper names (such as *the poet*, *Leopold*, etc.) are neither anaphors nor pronominals; personal pronouns (e.g. *she*, *him*, *your*) are [–anaphor, +pronominal]; reflexives, (e.g. *herself*, *themselves*, *each other*, *one another*) are [+anaphor, –pronominal].

Given these traits, three principles apply, resulting in obligatory coreference or non-coreference, or in optional coreference (Reinhart, 1983). These are given below with sentences to exemplify their predictions:

1) Three principles of binding theory:
    a. An anaphor is bound in a local domain. As pointed out above, here the term anaphor only includes reflexives and reciprocals. The local domain roughly corresponds to the phrase the anaphor is in. A reflexive then has to find its antecedent within its same phrase, as in (a); on the contrary, in (b) the antecedent in the local domain is Molly, whose gender does not agree with the reflexive (or, from another point of view, the antecedent is outside of the local domain):
        i. *she*$_1$ *wrote it herself*$_1$
        ii. **Leopold*$_1$ *asked Molly*$_2$ *to help himself*$_1$

b. A pronominal is free in a local domain. This means that, contrary to reflexives, pronouns must be not bound in their local domain, i.e. their antecedent has to be out of their clause, as seen in (c) but not in (d):

iii. *Leopold$_1$ asked Molly$_2$ to help him$_1$*

iv. **She$_1$ helped her$_1$*

c. A referential expression is free. The last principle posits that referring expressions refer independently, i.e. they do not link back to other expressions. This explains why (e), where Leopold is bound, is ungrammatical:

v. **He$_1$ asked Molly$_2$ to help Leopold$_1$*

The distribution of the of the three types of NPs in different configurations can be exemplified in the following table (adapted from Büring, 2005: 4):

	Anaphor	**Pronominal**	**Ref. expression**
No antecedent	*That it rains bothers himself$_1$	That it rains bothers him$_1$	That it rains bothers Leopold$_1$
Non-local ant.	*Molly$_1$ thinks he$_2$ loves herself$_1$	Molly$_1$ thinks he$_2$ loves her$_1$	*Molly$_1$ thinks he$_2$ loves Molly$_1$
Local antecedent	Leopold$_1$watches himself$_1$ in the mirror	*Leopold$_1$ watches him$_1$ in the mirror	*Leopold$_1$ watches Leopold$_1$ in the mirror

So, for example, the "anaphor × no antecedent" cell is not grammatical because the anaphor *himself* must be bound, and the "referring expression × local antecedent" cell is not grammatical because the referring expression *Leopold* cannot be bound.

Exceptions to the principles have been noted and motivated in various ways (e.g. Jackendoff, 1972; Pollard and Sag, 1992). Many examples are problematic in that grammaticality is not a binary, but rather a gradient. Some grammatical choices of referring expressions, as shown in the introduction (Chapter 1), are not ideal or felicitous but nonetheless perfectly grammatical. Moreover, this account doesn't distinguish between types of antecedents,

such as entities versus events, while it seems plausible that these should show differences in processing. If anything, event antecedents are much less common (see Chapter 5) than entities, and this has been linked to a higher saliency: infrequent antecedents are easier to access and process than frequent antecedents because they attract more attention (e.g. Pynte and Colonna, 2000; van Gompel and Majid, 2004).

The Gricean programme, stemming from the 1967 William James Lectures and later published in 1989, unlike Government and Binding is not structural but pragmatic in nature. While Binding theory aims to describe an idealised form of language underlying all natural languages, Universal Grammar, Grice's (1989: 4) primary interest is the difference "between meaning and use" and "to distinguish between the case in which an utterance is inappropriate because it is false [...] and the case in which it is inappropriate for reasons of a different kind".

Another important difference between Government and Binding and Grice, on a formal level, is thus the non-monotonicity of the inferences sprouting from meaning, which can be defeasible by the context or by new information. In the framework, the interpreter of an utterance needs to identify a) the referents involved, b) deictic parameters like space and time, and c) which of the possible meanings apply to the specific context in which the utterance was produced (Grice, 1989: 25). Pragmatic enrichment disambiguates the interpretation of an utterance, enriching and specifying the semantic meaning, thus bringing context into the interpretation of the parameters above.

The Gricean principles are reported in (2):

2) Cooperative Principle: Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.
    a. Quantity: 1. Make your contribution as informative as is required; 2. Do not make your contribution more informative than is required.
    b. Quality: Try to make your contribution one that is true. 1. Do not say what you believe to be false; 2. Do not say that for which you lack adequate evidence.

c. Relation: Be relevant.

d. Manner: Be perspicuous. 1. Avoid obscurity of expression; 2. Avoid ambiguity; 3. Be brief (avoid unnecessary prolixity); 4. Be orderly.

These maxims can be obviously related to the notion of prominence as outlined above: for example, the maxim of Relation (2c) could be used as a starting point to explain the fact that completely non-prominent (e.g. not at all activated in short term memory or even not available) referents, which require less economical forms to be reactivated, will not be brought up in a conversation unless there's a clear reason to. The maxims of Quantity (2a) and Manner (2d) may speak to the fact that a non-prominent referent may need stronger referring expressions to be brought forward in the discourse, so as to give enough information with a richer expression and not be ambiguous – or, conversely, naming a very prominent antecedent with an expression that is too rich would give too much information and not be brief. An interpreter would consider these violations of the maxims to happen purposefully (being guided in this reasoning by the Cooperative Principle in (2), and interpret the utterance accordingly.

2.2 Neo-Gricean accounts

Some Pragmatics theories stemming from Grice's account chose to reduce the model. In Huang's (1991, 2000a,b) and Levinson's (1987, 1991, 1995, 2000) neo-Gricean accounts, referring expressions are disposed in Horn Scales (Horn, 1972; see also Huang, 2000a: 208, and Gazdar, 1979: 55ff) and the maxims are reduced to three heuristics of Quantity, Informativeness and Manner: the Quantity principle lower-binds the set, so that if a weaker expression is used, the stronger expression's interpretation will be prevented, and the Informativeness principle upper-binds it, granting that to a general expression will follow a stereotypical and maximally enriching interpretation (while more complex referring expressions would override the stereotypical interpretation). The Manner principle is metalinguistic in nature, and states that producing a high-rank expression will implicate the negation of the interpretation of the lower expressions.

Referring expressions are ordered by their semantic richness and phonetic content: so, for example, the expression "the man" would be richer in content and phonetically weightier

than the pronoun "he." Both the semantic and phonetic load of an expression account for its "heaviness." Human language strives to be economical and yield the biggest referential precision with the least effort (see e.g. Haiman, 1985), and thus this concept of heaviness will be relevant, in one way or another, in most theories trying to explain the choice of antecedents and referring expressions.

Keeping in mind an updated version of Farmer and Hamish's (1987) Disjoint Reference Principle, stating that the co-arguments of a predicate will be disjoint unless provided with a reflexive marking, the general anaphora resolution strategy of Neo Gricean theories can be summarised as the synergy of two Manner and Informativeness implicatures (3):

3) lexical NP > pronoun > NP gap
   ← M-implicates non-coreference
   I-implicates coreference →

This strategy would predict that, in a language like Italian or Spanish, null pronouns will maximally implicate coreference to the most salient antecedent, while NPs would maximally implicate non-coreference. A similar prediction is broadly found in most theories of coreference resolution (e.g. Ariel, 1990; Carminati, 2002).

Another theory from the Gricean umbrella, Relevance theory (Sperber and Wilson, 1995, 2002 for a review) follows from similar premises but, from a cognitive standpoint, it reaches opposite predictions: while in Neo Gricean pragmatics the default is a maximally enriched generalisation, which can be prevented by world knowledge or by stronger referring expressions (at a processing cost due to the need to delete a "generalised conversational implicature": Levinson, 2000), Relevance theory does not adhere to this idea of default interpretive heuristics, stating that nonce implicatures only arise (at a processing cost) when they are needed to find relevance in a given context, every time afresh.

Wilson and Sperber (2004) start from Grice's Relation maxim (see 2c above) and develop the eponymous theory from both a cognitive and a communicative point of view: the cognitive principle then states that human cognition is geared to maximise relevance, and the communicative principle that utterances create an expectation of maximal relevance.

Both the neo-Gricean and the Relevance frameworks have called for experimental testing of their cognitive predictions (Levinson, 2000: 370; Sperber and Wilson, 2002: 6), and the evidence supporting or disproving them has been equivocal.[2] In any case, the problem is increased by the uneven definitions of accessibility and salience and prominence.

2.3 Accessibility ~ reduction models

Ariel's Accessibility Theory (1990) starts from the premise of context as seen in Sperber and Wilson (1982, 1995): a set of information that is, to a certain degree, available to the addressee of an utterance, allowing them to process a piece of information based on its accessibility. The theory, then, focusses on accessibility, and specifically on how NPs antecedents are accessed. It is worth noting that Ariel (1994) explicitly compares Accessibility Theory with Levinson's neo-Gricean framework. However, neither Ariel nor the neo-Gricean approaches (nor Relevance Theory) describe the psychological processes: memory and more general cognitive processes are, on the contrary, taken into account in Gernsbacher's (1990) Structure Building Framework (see §2.4).

In Accessibility Theory, the role of accessibility is directly linked to memory. To Chafe's (1976) and Gundel et al. (1993)'s question on how long the givenness (i.e. activation) of a referent lasts and how it is organised in a scale, Ariel adds that of how easy or difficult it is to access some material after it is no longer given. Reinstatement processes are necessary to retrieve information that is no longer currently activated in memory and make sense of a sentence which may refer to a previous point in discourse. Through the use of different referring expressions, then, a speaker signals how difficult the retrieval of a referent will be: different kinds of referring expressions give instructions about how to search for the referent in memory, and signal the different cost involved in the process. For example, in a trade off, an NP signals a higher processing cost involved in the search than a pronoun, but it also gives the interpreter more semantic information to aid its retrieval.

[2] For example, see studies on the implicatures produced by the word *some*: a) in favour of Relevance theory: Breheny, Katsos and Williams (2006); Noveck and Posada (2003); Huang and Snedeker (2009); b) in favour of the neo-Gricean framework: Grodner et al. (2010); Slabakova (2010).

For Accessibility Theory, then, referring expressions are "context retrievers", and the harder an antecedent (along with its contextual meaning) is to retrieve, the heavier their content must be. Highly accessible entities require a low-cost low-information referring expression, while not-very-accessible antecedents require high-cost high-information referring expression. Ariel (1990: 73) reports a hierarchy of referring expressions ranging from lowest to highest accessibility, a simplified version of which is given in (4):

4) Definite description — LOW ACCESSIBILITY
   Proper name
   Demonstrative
   Stressed pronoun
   Unstressed pronoun
   Cliticised pronoun
   Gaps, pro, PRO, wh-traces, agreement — HIGH ACCESSIBILITY

In a similar fashion, Givón (1983: 17) proposes a hierarchy which also includes syntactic structures, where more complex structures such as clefts and focussing constructions are used to code "most inaccessible topics". Cleft sentences as a marker of focus will be tested in Chapter 4, where clefts of the subject or the object are used to manipulate the information structural status (and thus the prominence) of a referent.

Accessibility levels change during a discourse: accessibility decreases when the distance between antecedent and anaphor increases. Moreover, when the number of competitors increases the accessibility of referents goes down: this happens even if an antecedent is not explicitly mentioned, as shown in Chapter 5. Other factors influencing prominence are grammatical function, animacy and the order of mention of a string of referents (with a first mention being generally more prominent).

Moreover, the information structural role of the entity influences its accessibility, with topics being more salient than non-topics. A non-topic like focus, for example, marks a referent as prominent for the speaker but not for the hearer, for whom it may be contrastive or still unknown information (e.g. Lambrecht, 1994). Its effect is more thus bound to

be more nuanced: while not clearly outlined in Ariel (1990), it seems to decrease the accessibility of a referent only insofar as its election as a topic in following discourse is concerned, but not with regard to referring expressions, that are on the other hand lighter (likely because of the increased prominence that focus lends to a referent: cf. Chapter 4).

More phonological or semantic material will be used to code a referent that has lost accessibility in discourse (Ariel 2004; Givón, 1989: 217), and conversely the more complex or heavier a referent is, the more mental effort will be required to process it (Givón, 1985: 197). In languages with rich verbal agreement, like Italian, the inflection of the verb may form an intermediate accessibility marker, helping retrieve the traits that could be encoded by the pronoun (Ariel, 1999: 226).

These prominence scales can be applied to contexts other than two stereotypical entity referents competing for coreference. Linguistic phenomena that happen more frequently, like reference to entities, will be cognitively more activated: reference to events, however, is rarer, and heavier referring expressions will be needed to perform it. An example of this is the gradable split whereby personal pronouns are used more to refer to entities, while demonstratives are used more for event antecedents (see Chapter 5).

At the same time, less common referring expressions may be used to indicate a highly prominent referent. This will be dealt with in Chapter 3 looking at Twitter language, a genre situated between spoken and written language (Koch and Oesterreicher, 1985). In the study, uncommon variants of some features, such as pronoun case or the use of graphic markers, are found to be stacked on top of each other to heavily express prominence.

Arnold (2010) reviews the claims on referring expressions and accessibility, with a focus on production. The two main constraints are found to be discourse status and processing constraint, with speakers choosing referential explicitness to help the listener. These constraints – along with Accessibility Theory's hierarchy in (4) – are seen as "prominence-lending cues" by Von Heusinger and Schumacher (2019), who show how the concepts of activation, accessibility and salience can be derived from that of discourse prominence. Prominence is characterised as a "structure-binding principle": it is relational, relative, it shifts

in time, and it can attract structures that would be less appropriate in non-prominent scenarios, such as the use of a certain referring expression, or the use of a cleft. This will be indirectly tested in Chapter 3, wherein a corpus study is used to explore whether different markers of emphasis are used independently or concomitantly. The use of a marked structure such as clefts is shown to overlap with that of other graphic markers of emphasis (or prominence), such as capitalisation or nonstandard punctuation.

The concept of prominence coming up in many theories on reference and anaphora is especially relevant to theories of pronoun interpretation. The assumption for languages whose pronoun inventory includes more than one subject pronoun is that the weaker form will be used when the entity to retrieve is more relevant – or accessible, or prominent – and the stronger form will be used when the context does not provide enough information to retrieve the entity, or where it may mislead towards another more expected referent. Thus in English, for example, a pronoun can be stressed to retrieve an entity that is less prominent (e.g. Kameyama, 1999, Gundel et al., 1993; although some evidence counters this: Kehler et al., 2008, for example, claim that this placement of stress is ruled by coherence rather than pronominalisation). In a language allowing null subjects, this division of labour is assumed to exist between null and overt pronouns (see e.g. Carminati, 2002 for Italian).

The status of the referent and, consequently, which referring expression should be used to refer to it are the main concerns of Centering Theory. Centering Theory was developed to account for the use of different referring expressions and their interaction with the coherence of a discourse (Grosz et al., 1983, 1995). Its main interest is then attention and coherence in discourse, and specifically how coherence and salience manifest locally in discourse segments.

Each sentence has a unique backward-looking centre and a set of forward-looking centres. The backward-looking centre links the sentence to what precedes, and the forward-looking centre is a set of entities to which the discourse could link to in the future. The forward-looking centres are ordered by their prominence, and elected as the next backward-looking centre according to this hierarchy. The prominence of the forward-looking centre depends on its grammatical role: some other possible factors not included in Centering are

surface position, thematic role, information structure and pronominalisation. If the backward-looking centre of the current utterance is the same as the one of the previous (centre continuation), a pronoun should be used to express it. Moreover, when the current centre is not pronominalized, the use of a pronoun for other entities is constrained: this use is only possible if attention is shifted back to a previously centred entity.

The misuse of referring expressions leads to processing penalties. For example, using a pronoun for a newly centred entity would cause the interpreter to first incorrectly link the pronoun with the previously centred entity, and only later backtrack to interpret it correctly (Brennan et al., 1987; Brennan, 1995). The theory, although maintaining a computational perspective, is then akin to theories linking a reduction in referring expression to the predictability or accessibility of its antecedent: since the easiest option in discourse is to maintain the previous centre (subject and topic continuation across sentences), this centre will be recalled using a light form such as a pronoun. Chapter 4 presents a study in Italian that uses sentence boundary to manipulate the status of the referents, or, in Centering terms, to update the centres: with story continuation studies, participants were able to both choose who to talk about (i.e. the backward-looking centre), and how to talk about it, that is to choose any expression from the Italian inventory to refer to it. It can be assumed that writers would produce referring expressions that are appropriate to the backward-looking centre of their choosing, and this choice would vary with the sentence boundary, grammatical role and information structure.

Centering theory remains rather underspecified and cannot thus produce strict predictions. Both the backward-looking centre and the factors influencing the ranking within the forward-looking centre are not defined or lack in details. Poesio et al. (2004) summarise the ways in which subsequent research has specified these and other deficits of the original theory, and find that it is possible to instantiate the theory such that its claims are verified, but only partially, with further theoretical and experimental work needed. Further, Taboada (2008) analyses a corpus of spoken Spanish finding that referring expressions are correctly predicted by Centering except in cases in which a name is used rather than a pronoun: this is explained with spoken-language constraints and genre difference.

2.4 Cognitively-motivated approaches

While the approaches presented so far are more steadily rooted in linguistic theory and its application, other theories are explicitly founded on cognitive aspects of language production and processing.

The Structure Building Framework (Gernsbacher, 1990) stems from similar considerations that language comprehension, whose aim is to build a coherent mental structure of the information in a text, uses more general cognitive mechanisms which are related to memory. Mental structures are further subdivided in substructures, and they are composed of "memory cells" that get activated and reactivated by incoming information and enhance each other's activation by levels of similarity: information which is relevant to the discourse activates information at the periphery of its semantics. If information isn't coherent enough to be fit into a substructure, a new substructure is created, and the activation of the previous memory cells is suppressed because it is no longer important for the building of further structure.

Gernsbacher (1990: 138–141) considers three properties that discourse has in relation with referential accessibility and the explicitness of anaphors (5):

5) Properties of the Structure Building Framework:
    a. Referential distance: the strength of an anaphor increases with distance from its coreferent. New concepts suppress older concepts in a competition similar to that of Centering (§2.3), and it is more likely that newer concepts may be brought up again with the passing of time. With each re-mention of a referent, its activation increases;
    b. Topicality: reference to topical entities is faster to process and easier to access, and is expressed through less explicit anaphors. First mention has the advantage of enhancing a referent's activation more strongly and durably;
    c. Episode structure: speakers use more explicit expressions at the beginning of "episodes", as referents are hard to access. Between episode boundaries, re-activation is more difficult, because at that point in a

discourse it is likely for a topic shift to happen. In the absence of a topic shift, an old referent is reintroduced.

Note that these properties share a commonality with the idea that a speaker will make choices in their production of referring expression based on a cooperation towards the listener, keeping in mind what may be better understood.

The ideas around short and long term memory of the Structure Building Framework also underlie the Informational Load Hypothesis. This was first proposed by Almor (1999) to account for the use and resolution of noun phrase anaphors. The hypothesis explains the processing of anaphors as an optimization process. Cost is proportional to the amount of semantic information activated by the referring expression: this is different from associating cost with the type of referring expression because it takes the (textual and extra-textual) context of the text into account, including discourse function: e.g. adding new information or re-activating it. Additional cost must serve a function in discourse, otherwise it wouldn't be allowed. This also correlates to the length of an antecedent's referring expression, as adding material to the form expressing a referent makes it semantically and pragmatically richer (e.g. Hemforth et al., 2013).

Since working memory lets information decay over time, with lesser activation the semantic overlap between antecedent and anaphoric expression must be bigger, to more explicitly make a link between the two. It is plausible that complex antecedents such as events or named entities would be costlier in terms of processing, and thus use heavier referring expression. This assumption is tested for events in Chapter 5 and for named entities in chapter 6, showing that for both the elements influencing such choices are multiple, and cannot be abridged to simple information decay and cost-effective processing.

The Informational Load Hypothesis incorporates basic ideas from Grice's Maxims (2 above). In particular, needing a reason to justify new information being brought into the discourse (in terms of cost) can be explained in terms of the maxim of Quantity (be as informational as required: no more, no less). The informational load of an anaphor with respect to a given antecedent needs to have a functional justification in either aiding the identification and/or adding new info about it. In the same way, the maxims of Relation

and Manner can partly explain the discourse properties regarding referent activation in the Structure Building Framework: something that is not relevant is less likely to be re-mentioned, and the discourse will be built in an orderly fashion, avoiding obscure referring expressions unless the context grants their use.

The Informational Load Hypothesis gives rise to interesting predictions with regard to focussed referents. When an antecedent is focussed it is highly activated in working memory, and so anaphors relating to it may be easier to process. This includes NPs, both when the anaphor is informationally heavier (e.g. *A bird… the albatross*) and lighter (e.g. *An albatross… the bird*). On the contrary, heavy anaphors which do not add any information (e.g. *A bird… the bird*) may be easier to process when the antecedent is not focussed: in the case of a focussed antecedent, the high cost of the anaphor would not be justified. Focus, in fact, has been shown to increase the cognitive prominence of a referent, thus making its retrieval with a pronoun easier (Cowles et al., 2007). However, these predictions seem to contradict Centering Theory (§2.3), where an NP anaphor would always be disfavoured when its antecedent is at the centre of attention, no matter what its informational contribution is, as well as other studies showing that focussed referents are processed more slowly across the board (Cowles and Garnham, 2011). Chapter 4, with a story continuation study, checks whether full noun phrases are used at all to refer to a referent that was focussed in the previous sentence.

Gelormini-Lezama and Almor (2011, 2014) further developed these ideas with an eye to Carminati's (2002) Position of Antecedent Strategy (see below §2.5), whereby syntactic function relates to form, with subjects being in the most prominent position and thus referred to with a null pronoun. In self-paced reading experiments in Spanish, non-null referring expressions caused processing delays when referring to salient antecedents and when referring to the first antecedent in the sentence order. However this did not happen marking a focus with a cleft, where the context is emphatic. These results go counter to Carminati's prediction that non-null expressions would always be inappropriate for subjects (including an emphatic focussed subject).

2.5 Newer frameworks

Finally, some frameworks strive to expand the work on coreference beyond the usual assumptions: Carminati's (2002) Position of Antecedent Strategy focusses on Italian studying how a different pronominal repertoire, which includes both overt and null pronouns, differ from the widely studied English patterns, while Bayesian approaches (e.g. Kehler et al., 2008) model the difference between interpretation and production of anaphoric expressions in probabilistic terms.

Carminati's (2002) work starts from a generative syntactic background and focusses on Italian, rather than English, but was later applied to multiple (Romance) languages. Her Position of Antecedent Strategy (PAS) stems from Centering Theory (§2.3) and it assumes the existence of a division of labour between Italian null and overt pronouns, wherein null pronouns used anaphorically refer to the most prominent antecedent (6a), while overt pronouns refer to less prominent antecedents (6b):

6) Carminati's division of labour:
    a. Marta$_1$ ha visto Sara$_2$ quando Ø$_1$ andava al mercato.
       Marta$_1$ saw Sara$_2$ when Ø$_1$ was going to the market.
    b. Marta$_1$ ha visto Sara$_2$ quando lei$_2$ andava al mercato.
       Marta$_1$ saw Sara$_2$ when she$_2$ was going to the market.

In the theory, prominence is defined syntactically, with the SpecIP (preverbal subject) position being the most prominent. The notion of relative prominence is influenced by Centering Theory: in the PAS only structural factors have an influence on prominence. Carminati's research only looks at pronominal anaphors, citing Ariel (1990; see above §2.3) saying that heavier referring expressions (like nouns, demonstratives, etc.) should be used to refer to even less salient referents than the overt pronoun. However, it is unclear how this would play out within the division of labour between null and overt pronoun.

It is worth noting, moreover, that the Spec IP position is more prominent than higher positions that may be filled in the CP: this implies that the subject antecedent will be the most prominent and retrieved by null pronouns even in the case of structures such as left

dislocations or clefts, marking respectively a topic and a contrastive focus on the moved antecedent.

Another prediction that can be gathered from this theory is the fact that semantic or world knowledge factors (e.g. the meaning of the verb, the need to disambiguate between similar referents, other factors weighing on working memory and processing like the distance and number of the antecedents, etc.) will not have a role in choosing a referring expression.

The PAS has been applied and tested psycholinguistically in Italian (e.g. Filiaci, 2010), in Spanish (Filiaci, 2010; Alonso-Ovalle, Fernández-Solera, Frazier and Clifton, 2002), Catalan (Perera and Bel, 2011), Portuguese (Baumann et al., 2011) and Romanian (Geber, 2006), all languages allowing a null subject pronoun like Italian, but the support the theory's predictions get from these studies is inconsistent. For example, Filiaci (2010) finds a processing penalty when an overt pronoun is resolved against its bias only in Italian but not in Spanish, and Alonso-Ovalle et al. (2002) find that the PAS is overridden in certain syntactic contexts in Spanish (namely when overt pronouns are pre-verbal in embedded clauses).

As can be seen from the attempt to extend Carminati's account to other languages (but also in the other accounts mentioned so far), it can be hard to apply rules or heuristics devised for one language to another. An important factor in this when it comes to referring expressions and coreference is the availability of null pronouns, but even with languages similar from a typological point of view, some generalisations are found to not be appropriate (e.g. Spanish and Italian pronouns seem to behave differently, despite both languages' null-subject option: see e.g. Filiaci et al., 2014).

In null-subject (or pro-drop) languages, null pronouns add a further level of minimality to the set of referring expressions compared to English, which is still the dominant starting point for linguistic theory. This causes a shift in the application of base-rules such as Grice's (1989) maxims: when the time comes to apply the second maxim of quantity, for example because reference is so obvious that minimal help is needed on the part of the speaker to

identify it, the minimum amount of informativity will be given by a null pronoun rather than the usual, prosodically unmarked subject pronoun.

This leads both to English-based theories being inaccurate on null subject languages such as Italian or Spanish (lest a shift is explicitly made, equating the English pronoun to a null pronoun and so forth), or to theories made on a null subject language to focus on the duality of null vs overt pronoun disregarding the array of all possible referring expressions (like in Carminati's PAS, where a division of labour is set up between the two disregarding other types of pronouns or nouns).

A further assumption often taken for granted in the literature about pronoun interpretation is that the production side of pronouns – the speaker's job, rather than the listener's – mirrors the interpretation side, following the same set of rules or biases. This is the image of an efficient and perfectly collaborative world, in which a speaker will use a certain referring expression only if they believe this to be the mark that will be correctly interpreted by the listener, and the listener, at their end, will interpret the expression with the firm belief that the speaker made all efforts to be flawlessly cooperative. However, psycholinguistic studies have come to challenge this idea of the speaker as a stereotypically collaborative participant in some areas of language (e.g. Bonnefon et al., 2009; Fukumura and van Gompel, 2012; Spector, 2014). This can be seen in coreferential patterns too.

Stemming from Gricean ideas, Probabilistic Pragmatics (see e.g. Franke and Jäger, 2016) models pragmatic phenomena with a Bayesian approach. This is a cluster of approaches deriving from psycholinguistics, pragmatics, game theory and decision theory, all aiming to explain linguistic phenomena in pragmatics "as rational or optimally adapted for a conversational purpose" (37).

Rather than defining interpretation as simply the mirror image of production (or vice versa), Probabilistic Pragmatics models it as a reverse-production through Bayes's Rule. This is meant to describe regularities in the language use of speakers and comprehenders as practice meant to make a conversation possible despite the uncertainty in meaning and anaphoric relations. Kehler et al. (2008) relate this to pronoun resolution applying to it the equation in (7):

7) $P(referent \mid pronoun) \sim P(pronoun \mid referent)\ P(referent)$

where the right hand side is the productive side wherein P(referent) is the prior probability that a speaker will mention a referent next and P(pronoun|referent) is the conditional probability that given a speaker is mentioning a particular referent, they will use a pronoun as a referring expression (the likelihood); the left hand side, P(referent|pronoun), on the other hand, is the comprehender bias, the probability that given a pronoun, a referent will be chosen as its interpretation. As per Bayes' Rule, the two sides are proportional.

This model can take two interpretations: a Weak one in which pronoun interpretation and production simply stand in a Bayesian relation with each other, and a Strong form in which the prior and the likelihood are influenced by different factors. What this means is that in the Strong claim pronoun production (the prior) is posited to be sensitive to structural factors whereas pronoun interpretation is posited to also be sensitive to meaning. Next-mention bias, for example, would be influenced by verb semantics and coherence relations: this can be seen clearly with verbs of Implicit Causality, in which a verb biases the prominence of the referents involved in its structure biasing toward the selection of one of them. This semantic feature biases the selection of the referent, but not the choice of referring expression to use to talk about it (Fukumura and van Gompel, 2010). Information structure, on the other hand, would influence the production of referring expressions but not antecedent selection (Rohde and Kehler, 2014).

The model has been put to the test in Portuguese in an experiment by Fernandes et al. (2018), where pronoun resolution was shown to be influenced by the overall probability of a pronoun, while keeping in line with models wherein the prominence of a referent is associated with lightness of referring expression. An example of information structure influencing the choice of referring expression, conversely, can be seen in the experiments of Colonna et al. (2012) and De la Fuente and Hemforth (2013) among others (see Chapter 4), even if they are not directly related to Bayesian models: in the interpretation of intra-sentential pronouns in German and French, topicalisation enhances, but focussing reduces the accessibility of antecedents, with effects on the interpretation of referring expres-

sions. The two studies conclude that while topicalisation enhances prominence, focalisation hinders it. Other studies show how anaphors to non-focussed antecedents are read more quickly (Cowles and Garnham, 2011). However this last study, testing the Informational Load Hypothesis (Almor 1999) psycholinguistically, did not manage to replicate the effect in two subsequent studies. Moreover, evidence has also been found of focalisation increasing prominence (Cowles et al., 2007; Klin et al., 2004).

2.6 Overview of the next chapters

In the four chapters following, six studies are presented and their results compared to the theoretical predictions. They are organised in two sections respectively on the surface form of prominence and its repercussions on antecedent and referring expression selection, and on prominence as seen in different components of meaning, specifically in event structure and named entities.

These components lend themselves to testing the theories outlined above from different perspectives. For example, models predicting the reduction of a form from the prominence of its referent can be tested looking at how prominence, as encoded by information structure, influences the choice a writer makes (Chapter 4), as well as by looking at how complex events are encoded when compared to more simple entities (Chapter 5).

Simultaneously, the cross-linguistic approach of the thesis gives an opportunity to test whether some general theories usually modelled on English can be extended to other languages. In particular, Chapter 3 is on English, Chapter 4 on Italian, and Chapters 5 and 6 test five languages: English, French, German, Italian and Spanish. These are typologically similar enough to make the application of English-based theories plausible, but still variable in some fundamental features, such as the availability of a null subject.

All in all, this thesis takes an approach that attempts to look beyond some common assumptions: the ideas that interpretation and production mirror each other, that opposite forms are simply distributed in functional divisions of labour, and that general theories designed starting from English can be overextended from language-specific to language-general.

Every chapter will include both a small literature review relevant to its topic, and empirical studies, as well as an introduction and conclusion to situate them in the broader scope of the thesis. Some of the chapters (namely chapters 3, 5 and 6) include published or currently-in-review articles, some of which are co-authored. Details on authorship and contribution to each piece of research are given as a preamble to the chapters.

Chapter 3 The surface form of prominence

How do writers express prominence through form?

3.1 Overview

Before looking at how prominence works and how it affects antecedent choice and referring expressions, this chapter only looks at its surface form in written English to explore how prominence is conveyed, whether there are multiple ways to signal it and if these modes are used interchangeably or concomitantly. In particular, the study targets cleft sentences, marked structures known to signal some kind of prominence in their own right (e.g. Lambrecht, 1994; see §3.2.2 below), in two different genres/media of written English: Twitter and iWeb. While iWeb (Davies, 2018) is a corpus of edited, standard written English gathered by sourcing texts from a curated collection of websites, the status of Twitter language is more ambiguous: it is to begin with produced by individuals in the context of social media, and thus not curated according to specific editorial choices, nor is it a standard variety of English, seeming to sit somewhere in between spoken and written language, in the middle of the conceptual orality gradient (Koch and Oesterreicher, 1985). This concept describes how the medium and conception of a text can go in different directions: an utterance can be realised as spoken or written in the medial dimension, but be conceptually at different points between these two poles. Twitter language, as an example of computer-mediated communication in the context of social media, is in a written medium but conceptually more leaning towards orality.

This study focusses on clefts, a structure expressing the prominence and, more specifically, contrastive emphasis of a referent. Within this construction, various elements show variation in form, and these may correlate to how a speaker may want to express the prominence of some content. In example (1), the cleft focusses the pronoun "him" as contrastive or new information. As can be seen, clefts in English include a pronoun "it" followed by a copula, the clefted referent, a complementiser introducing a relative clause expressing the content of the corresponding unmarked sentence (in this case, *He should be put in a straight jacket*):

1) it's HIM who should be put in a straight jacket!

The "formal" features of the clefts analysed in the study include the use of emphasis markers, that is graphic ways to signal a marked intonation such as ALL CAPS, *asterisks* and so on, as well as variants of three components of cleft sentences: whether a contraction is present or not at the beginning of the cleft (*it's* vs *it is*), the case of the clefted pronoun (*she* vs *her*) and the choice of complementiser (*that* vs *who*). Moreover, the syntactic nature of the clefted element (subject vs object) is considered. Notice that the case of the pronoun does not necessarily encode the syntactic role of the clefted element (although that is one of the possibilities), as shown in the example in (1) above, extracted from the Twitter data used in the study: the case of the pronoun could thus be refunctionalised to signal something else, like prominence. Example (1) shows an accusative pronoun used to cleft the subject of the relative clause (also called *presupposed clause*; see Prince, 1978) and marked as emphatic through capitalisation (and an exclamation mark at the end of the sentence, a perfectly standard emphasis marking), as well as the use of contraction and *who* as the chosen complementiser. As some of these features are clearly liable to vary in distribution with the genre and formality of texts – e.g., it seems unlikely that emphasis markers would be used in formal registers, and the other features may have a preferred or prescribed formal variant –, the study also compares the form of clefts in Twitter and edited written English.

The findings of the study are partially exemplified by this sentence: emphasis markers can indeed be used within clefts, thus "stacking up" markings of prominence. Moreover, as shown in (1), the case of the pronoun is not chosen to be in line with the syntactic role of the referent. Finally, the three variants are used to mark prominence with similar expectations to those in "accessibility ~ reduction" models of coreference: while not being as directly influenced by the prominence of a specific referent, here too variants that are phonologically (*it is* rather than *it's*) or semantically stronger (*who* rather than *that*) are used concurrently with other prominence markings, and in the case of pronoun case the variant which marks more prominence is the rarer one, that is, the nominative, also perceived as the more formal variant.

Finally, in the comparison between Twitter and the standard English used in edited writing in iWeb, some differences are worth noting. Twitter English clefts are contracted much more often than edited writing's, and accusative pronouns appear much more frequently. This too can be explained as a vicinity of Twitter language to the spoken genre in spite of its written medium.

The research presented in this chapter was undertaken in the context of a fellowship at Universität Potsdam and funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Project ID 317633480 – SFB 1287, Project A03/MGK. The study was designed, conducted and analysed by myself under the supervision of Tatjana Scheffler and the rest of project A03. Tatjana Scheffler also contributed to writing the paper and researching the literature.[3]

The article, published in Linguistics Vanguard, is presented verbatim with the exception of section, figure and table numbering. In the article, the data set as well as the R code of the analysis are referenced in Note 10 as supplementary materials. These can be found at the article location, linked in Note 4.

3.2 Form variation of pronominal it-clefts in written English[4]

Clefts are well-studied as a construction which induces emphasis on its clefted referent. However, little is known about the distribution of different stylistic forms of it-cleft variants.

We report on a corpus study mining data from Twitter, targeting sentences clefting a pronoun in English. We examine the following features: case and syntactic role of the clefted pronoun, contraction of the copula, choice of complementiser and use of emphasis markers. The results show systematic associations between these features. A further compari-

[3] Also see Acknowledgements.

[4] Bevacqua, L., & Scheffler, T. (2020). Form Variation of Pronominal It-Clefts in Written English. A corpus study in Twitter and iWeb. *Linguistics Vanguard*, *6*(1). Available at https://doi.org/10.1515/lingvan-2019-0066.

son between the Twitter dataset and data from iWeb, a corpus of general-use web language, shows significant differences in levels of emphasis and formality, positioning Twitter language in the middle of the conceptual orality spectrum.

3.2.1 Introduction

Cleft constructions such as (1)–(3) are used to focus the clefted referent, enhancing its prominence and making it more accessible for re-mention in subsequent discourse.

1) It's she that this post honors.[5]
2) It is him who has caused her to smile brightly.
3) What John lost was his keys. (Prince, 1978)

There are well-studied and clear pragmatic differences between English it-clefts such as (1–2), and *wh*-clefts (3) (Prince, 1978). In contrast, there is less existing research discussing possible differences between form variants: the use of contraction, case of the pronoun and complementiser vary between sentences (1) and (2). In this paper, we investigate these form variants in more detail. In particular, we want to determine what means are used to mark or emphasise the clefted referent. To do this, we study naturally occurring it-clefts in data extracted from the social media platform Twitter: this allows us to tap into relatively spontaneous utterances containing clefts as a focussing construction. We show that the observed variation is by no means random: instead, we identify formality as a driving factor for expressing emphasis in the Twitter clefts, which leads to three preferred variants of it-clefts in English spontaneous written text. We compare our findings with another corpus of web data, i.e. iWeb (Davies, 2018), containing more strongly edited but still varied content. This corpus shows much less variation, because standards promoted in editing guidelines result in an almost categorical preference of certain forms. Our systematic investigation of a relatively rare phenomenon in a big dataset opens up new directions for explaining expressions of focus, emphasis, and the formality gradient in English.

[5] All examples are attested sentences from Twitter, unless marked otherwise (see Section §3.2.3.1 for details on the data collection).

In the following section, we report some findings on it-clefts, emphasis, and formality from the literature (Section §3.2.2). Afterwards, we present our corpus study, first within Twitter, and then as a comparison between Twitter and iWeb (Section §3.2.3). We note a systematic association of more informal (contraction, accusative-case pronouns, *that* as a complementiser and emphasis markers) and formal (uncontracted clefts, nominative case, and *who*) features, as well as a general wider variation of these features in Twitter writing than in iWeb's, positioning Twitter in a more central position in the "conceptual orality" spectrum (Koch and Oesterreicher, 1985). Finally, we discuss these results in the light of the background literature and rules of prescriptive grammars (Sections §3.2.4 and §3.2.5).

3.2.2 Background

3.2.2.1 It-Clefts

If a focus is pragmatically needed, e.g. to contrast a piece of information, it will usually be grammatically marked, either prosodically or morphosyntactically, or both (Lambrecht, 1994, p. 64). The expression of this marking can eventually be conventionalised, and cleft constructions are a good example of such grammaticalisation.

English it-clefts consist of the pronoun *it*, followed by a form of the verb *to be*, a clefted constituent, and a complementiser, which introduces a relative clause that is attributed to the clefted phrase (Lambrecht, 2001). It-clefts introduce two meaning parts: (i) a presupposition that the property in the clause following the complementiser holds of some entity; and (ii) an assertion (at-issue contribution) that this property holds of the entity denoted by the clefted constituent. We will thus call the clause following the complementiser the *presupposed clause*, following (Prince, 1978); the clefted constituent is also called the focussed constituent.

The literature on clefts has concentrated on several central questions of syntax and semantics. On the one hand, research has addressed the issue of the underlying syntactic structure of cleft sentences. On the other hand, given that the truth conditions between a cleft sentence and its corresponding simple clause do not differ, there has been discussion

about the exact pragmatic contribution of using a cleft construction, especially with regard to information structure (Delin, 1992).

Here, we focus on the form of the cleft construction itself, which offers a lot of variability within the frame defined above. We identify the following "moving parts" of it-clefts:

4) a) **Contraction**: The copula *be* may appear contracted or uncontracted.
   b) **Case**: The clefted constituent may be in nominative or accusative case. In English, this feature is only visible in pronouns.
   c) **Complementiser**: The complementiser in the relative clause can be *that* or a *wh*-complementiser. In the case of clefted objects, the complementiser may even be empty (Quirk et al., 1985, p. 1385), but we did not consider this particular case.[6]

The example in (1) illustrates a contracted, nominative cleft with the complementiser *that*, whereas (2) is an uncontracted, accusative cleft with *who* as the complementiser.

Of these features of variation in clefts, Maier (2013, 2014) studies the case of the pronoun using corpora. Maier (2013) shows that the form of the pronoun is influenced not only by its position in the sentence, but also by its function, with variation related to the discourse medium: while written language prefers nominative-case pronouns, spoken English uses more accusatives. Maier (2014) accounts for the variation in case with information structural rationale, claiming that the nominative case of English pronouns in clefts was refunctionalised to signal emphasis (which he calls "focus"). We note that since he investigates both it-clefts and other related constructions together, the conclusion that it-clefts with nominal case pronouns are more emphatic than the accusative ones does not strictly follow from his statistical analysis.

The case of the pronoun in it-clefts is also mentioned by Patten (2010, 2012). Patten notes how the case of the pronoun is fundamental to the type of syntactic account: under an expletive account (in which the cleft is used to mark a sentence with the addition of a "dummy" element), the case of the postcopular element should be nominative, but this

[6] Object clefts are rare, see discussion below in Section §3.2.3.3.

pattern "is not found in the dialect of the majority of speakers" (Patten, 2010, p. 146, reporting a finding in Akmajian, 1970). However, using the ICE-GB corpus (Nelson et al., 1998), Patten (2012) shows that clefted pronouns in the nominative are almost twice as frequent as those in the accusative (with a proviso of very low counts for either), and agrees with Akmajian (1970) in interpreting this as the preference for a prestige form, which gives an appearance of formality. This is further confirmed by a corpus study in Biber et al. (1999), where clefts with a nominative pronoun and *who* are frequent in fiction and news despite a general predominance of the accusative form. This is motivated claiming that nominative forms may be perceived to be more correct, possibly because of prescriptive rules moulded on those of Latin grammar.

As for complementisers, Collins (1991) finds that about two-thirds of it-clefts contain *that*, but most other *wh*-words are viable for the role. Cheshire et al. (2013) conduct an in-depth investigation of complementisers in restrictive relative clauses. They find that *who* is used almost exclusively in subject relative clauses. The choice of complementiser shows a strong variation in their data, though for younger speakers, most instances are realised with *that*. The complementiser *that* is the most general (lacking the animacy requirement of *who*). For some speakers, Cheshire et al. (2013) conclude that the subject preference of *who* is being grammaticalised as (additional) "topic" marking, in the sense that referents relativised with *who* are more salient (therefore, more available for re-mention or more "topic-worthy") in subsequent discourse.

Looking at their distribution across registers, Collins (1991) notes that clefts outnumber pseudo-clefts in writing, while the opposite happens in speaking. Within spoken genres, it-clefts are more used in prepared monologues than in dialogues, further pointing in a direction of higher formality. However, internal features such as the case of the pronoun or the complementiser of the cleft are not analysed.

We are not aware of studies that systematically investigate multiple axes of variability in English it-clefts. In this study, we will consider these variants within a theory of emphasis marking.

3.2.2.2 Emphasis

We define emphasis as a linguistic means chosen by a speaker to direct a hearer's attention, for example to help single out an intended referent (Zimmermann, 2008). This concept has been inconsistently referred to throughout the pragmatics literature, being at times likened to prosodic stress (e.g. Heath, 2018) and often used interchangeably with other under-defined or theory-specific terms such as "prominence" or "salience". Clefts of all kinds are emphatic in the sense that, as reflected by their prosodic prominence (Lambrecht, 2000), they identify the clefted element as something that the speaker considers especially worthy of attention, because of its status as new or contrastive information. In addition to the grammaticalised emphatic function of clefts, syntactic or morphological (or indeed prosodic) cues exist and can be used by the speaker to indicate particular emphasis.

Computer-mediated communication, being written, does not mark emphasis prosodically. However, it makes available a large range of other emphasis markers, that have been shown to take information structural roles. McAteer (1992) studies how the alteration of typeface can mark information focus in a way completely akin to intonational signals. Some of these are standard in writing and have conventionalised functions, such as punctuation or paragraphing, others are not, like all caps or boldface. These functions are included by McAteer (1992), along with clefting, in a set of "foregrounding strategies" that a writer can employ to mark contrast and emphasis. In more recent work on writing in social media, Heath (2018) identifies non-standard capitalisation as a marker of focus and emphasis, while Scott (2015) analyses hashtags as topic markers.

To study the variable encoding of the cleft construction and its relation to emphasis marking, we annotate these types of emphasis phenomena in our data:

5) Types of emphasis marking:
    a) **Capital pronoun**: The clefted pronoun is capitalised:
       e.g. "it's HE that's threatened us"
    b) **Capital initial**: The first letter of the clefted pronoun is capitalised:
       e.g. "it is She who has to live with the consequences"

c) **Capital cleft**: The whole cleft is capitalised:
e.g. “it’s HIM WHO IS THE PROBLEM”

d) **Exclamation**: The cleft contains exclamation points:
e.g. “it is he who holds all the baggage!”

e) **Non-standard comma**: Non-standardly placed comma to mimic prosody:
e.g. “it is he, that is my one true love”

f) **Asterisks**: Asterisks around the clefted pronoun:
e.g. “it’s *he* who did it”

g) **Other**: Other strategies, such as tildes or inverted commas around the clefted pronoun.

Social media has been shown to innovate written language systematically, with a big influence coming from spoken language: for example, spelling on Twitter appears to follow phonological variations relatively frequently (Eisenstein, 2013a). Non-standard emphasis markers are, likewise, more widespread in social media than in standard written language.

3.2.2.3 Formality

Another dimension which interacts with emphasis, in particular in social media, is formality. Heylighen and Dewaele (1999) define expressivity of language (which is another reshaping of emphasis in that it corresponds to linguistic genres with high detail and emotionality) as well as formality in terms of the orthogonal variability axes fuzziness and context-dependence. In this view, context-dependence refers to the amount of inference that has to be made in order to interpret an expression, fuzziness refers to the opposite of semantic precision of expression. Formal style will tend to minimise fuzziness and context-dependence, while expressive language is precise (not fuzzy) but highly context-dependent.

In this account, expressivity is a linear combination of precision (–fuzziness) and context-dependence: for example, a proper name will be much less fuzzy than a pronoun, which will in turn be highly context-dependent. Likewise, the two complementisers will vary

along the fuzziness axis: *who*, with its [+animate] trait, includes more information than *that*, and being more precise it will be more formal and more expressive. Its use could thus correlate with the avoidance of contraction in the cleft sentence (which increases its phonological weight and thus the formality of the register: see Kjellmer, 1997, for extensive examples). Similarly, the use of emphasis markers indicates increased expressiveness and could thus be associated with uncontracted clefts, or with a more precise form of the complementiser like *who*.

Further, it is known that social media are a highly informal medium, often reflecting features frequently found in spoken conversations more than written texts (cf. Scheffler, 2017; Storrer, 2013; Tagliamonte and Denis, 2008). This places social media texts in the middle of the "conceptual orality" spectrum (Koch and Oesterreicher, 1985) which moves along an axis of medium (spoken to written) and one of formality (informal to formal). Since emphasis by itself is also an effect of markedness, the specific use of a formal variant in social media would constitute a marked case and thus increase the overall emphasis of the utterance. In the following, we investigate the use of formal and emphatic variants of it-clefts in Twitter data.

3.2.3 Empirical study

We present a systematic empirical study which aims to classify it-cleft variants along emphasis and formality dimensions. It-clefts are a rare phenomenon in spoken and written corpora: for example, the iWeb corpus (Davies, 2018) only contains about 4000 of them, a very low number for a corpus including texts from more than 22 million web pages. In addition, a well-founded comparison of naturally occurring it-clefts from corpora should aim to exclude factors which can potentially confound the information structure, syntactic variation, and interpretation of the utterance. In particular, this concerns the syntactic and prosodic type of the clefted phrase, as well as its meaning (for example, length and weight of the phrase, animacy of the referent, etc.). We therefore choose to look only at it-clefts with clefted third-person pronouns. Unfortunately, this restriction makes most existing corpora too small for statistical analysis. Mining Twitter for data, conversely, still provides a good amount of examples to study. In addition to data availability, Twitter as a medium

has another advantage for studying this phenomenon: as indicated above, since "standard" language norms are less applicable here, we expect increased variability in the observed forms of structures such as clefts.

We first mined tweets with it-clefts and one of the authors annotated them (see Section §3.2.3.1). After an initial analysis, new Twitter data was collected to internally validate the findings using the same analysis on a second sample. Finally, we extracted clefts from the iWeb corpus (Davies, 2018), replicated the analysis in its data, and compared the results with the Twitter data.[7]

3.2.3.1 First data collection and annotation

Tweets including at least one cleft were collected on 25th to 29th March 2019 using TAGS (Hawksey, 2016). The queries were url-encoded to target specific tweets without excluding other factors such as emphasis markers or punctuation, and excluded retweets.[8] The features targeted in the searches are listed in (4): contraction, case of the pronoun, and complementiser. This means all it-clefts introduced with or without a contraction (*it's*[9] and *it is*), with a 3rd person singular pronoun in either nominative or accusative case, and *that* or *who* as a complementiser, followed by a presupposed clause of any length.

All tweets using a cleft in the context of a quote from sacred scriptures (e.g. "But thou shalt remember the LORD thy God: for it is he that giveth thee power to get wealth") were manually removed from the dataset because they are representative of a different register of language. Many of these appeared multiple times as "quotes of the day". This left the dataset with a total of 798 tweets that conform to our constraints.[10]

[7] We also obtained the English portion of the large multilingual spoken cleft corpus extracted from European Parliament recordings (Bouma et al., 2010). However, the number of clefts that match our criteria was too small to replicate our analysis directly.

[8] All queries are reported in Appendix A.

[9] We thank an anonymous reviewer for pointing out that the informal spelling variant *its*, lacking the apostrophe, is not matched by our queries, thus excluding this group of (plausibly) informal tweets from the data set. While this does potentially skew the overall view of our Twitter data towards increased formality, we chose not to include this variant as it is virtually absent from edited writing like that in the iWeb corpus, making the comparison of the two genres not possible.

[10] The raw numbers for each feature are reported in Appendix B. The full data set can be consulted in the supplementary materials to this article.

The data was manually annotated for the following features:

6) a) **Number of referents**: 1 for clefts wherein only one referent was mentioned (e.g. "it's he that needs to go"), 2 for clefts involving two referents (e.g. "it's she that called him").

   b) **Role of the clefted element**: Subject (e.g. "it is he that assisted Castro") versus Object (e.g. "it is he that I will respond to"). This annotation only concerns clefts with two referents (i.e. clefts with both a subject and an object).

   c) **Emphasis**: presence of emphasis markers: see Section §3.2.2.2.

   d) **Cleft**: text of the cleft(s), isolated from the rest of the tweet.

45 tweets (5.6% of the total) contained multiple clefts. Only the first cleft of every tweet was considered in the following analysis.

3.2.3.2 Twitter data analysis

The first analysis targets the full dataset and some of the features annotated in the tweets. The basic distribution of the three main factors listed in (4) (contraction, pronoun case and complementiser) is shown in Figure 1. It can be gathered that in general clefts in English-language tweets are more often contracted and contain an accusative pronoun with the complementiser *who*. Moreover, more clefts with masculine than with feminine pronouns are produced (489 vs 309, 61.3% vs 38.7%; this ratio is not visualised).

Figure 2 shows the distribution of all eight combinations of the main features considered. It seems evident that the features are not distributed randomly: three combinations dominate the picture. About equally frequent, they together account for about ¾ of the it-cleft occurrences. These are exemplified in (7):[11]

7) **Uncontr-NOM-who**: it is she who has the power
   Contr-ACC-who: it's her who has to do it
   Contr-ACC-that: it's him that is the target

[11] Note that all pronouns are syntactic subjects.

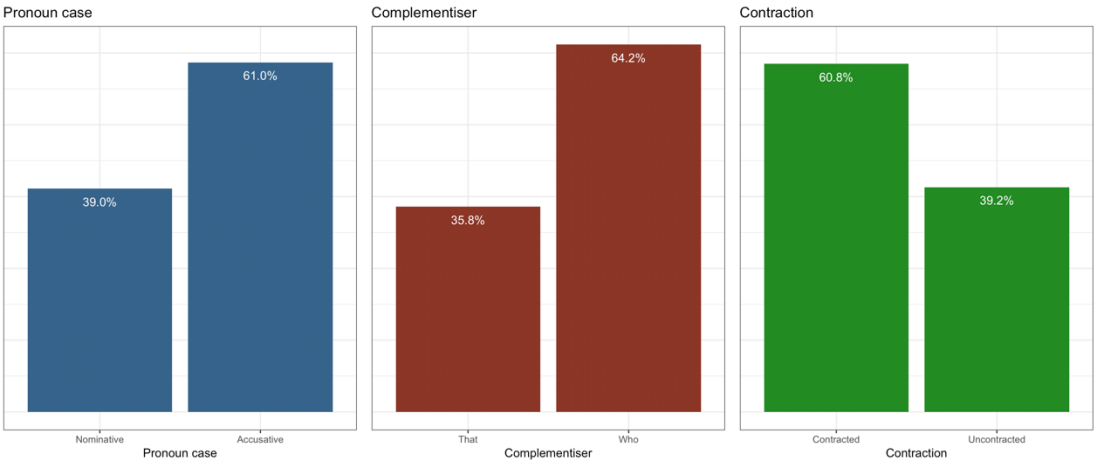

Figure 1. Ratios of the three features in the data set.

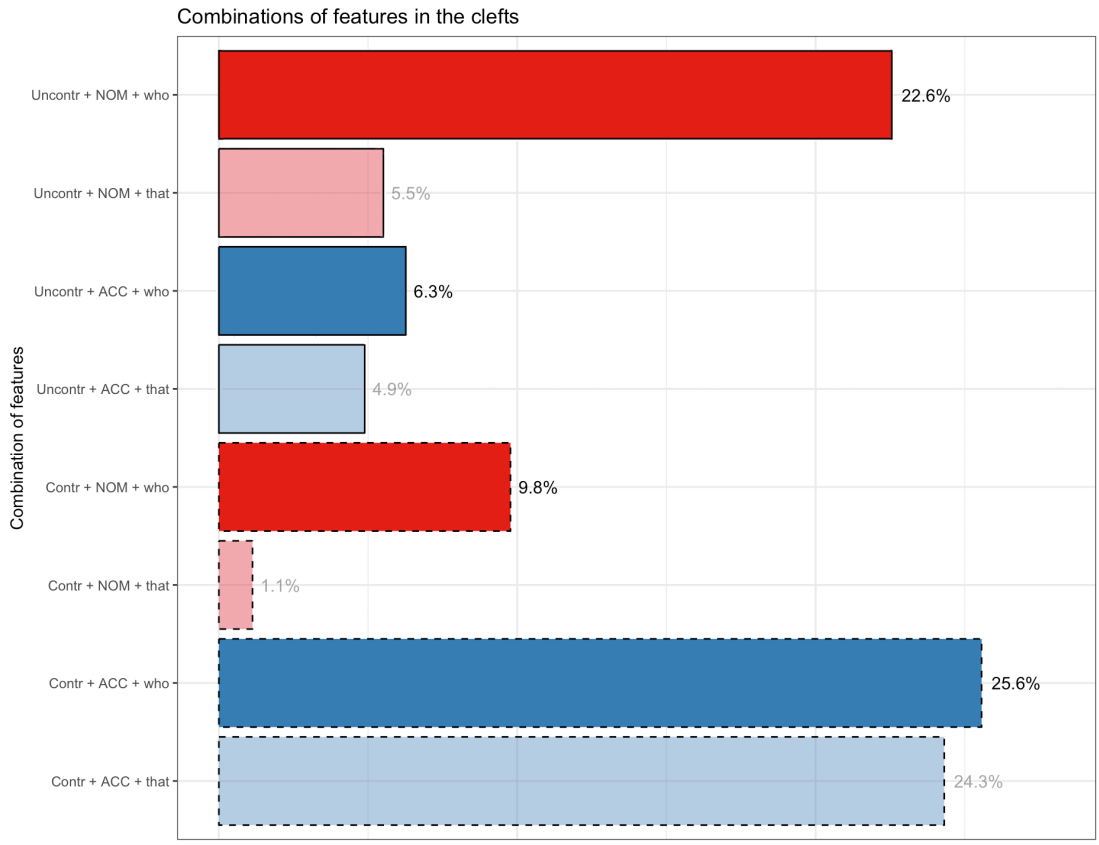

Figure 2. Distribution of the eight feature combinations in the data. The outline of the bars distinguishes uncontracted (continuous line) from contracted (segmented line) clefts, the colour of the bar distinguishes the case of the pronoun (nominative = red, accusative = blue), and the intensity of the colour distinguishes the complementiser (*who* = full colour, *that* = transparency).

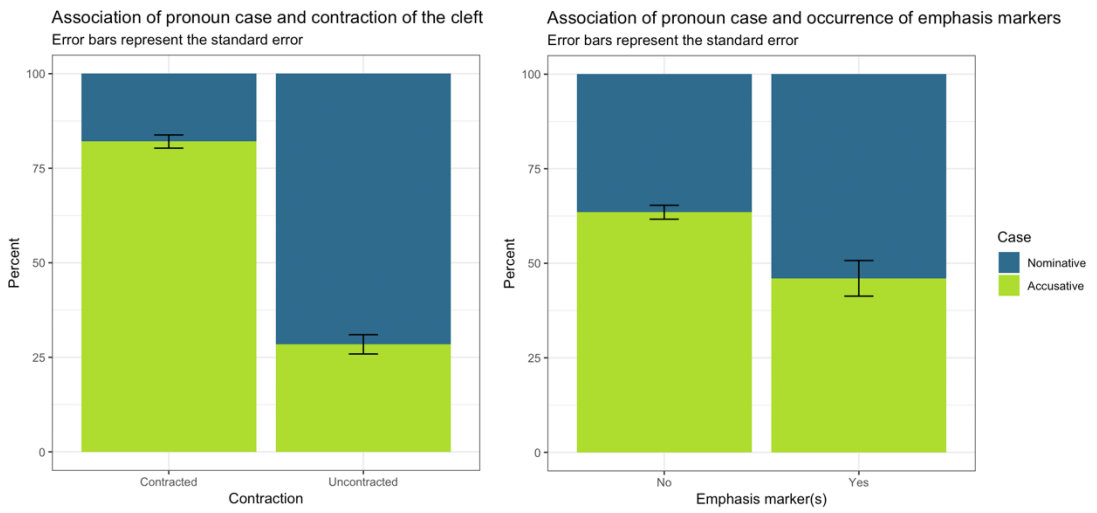

Figure 3. Associations of pronoun case with contraction and emphasis. Figures 3 to 6 show the co-occurrence of two features, represented on the x axis and through colour. The y axis shows the proportion of each combination.

In the following, we statistically evaluate the correlations between the individual features. Given that no direction of effect can be discerned, chi-squared tests[12] were used to test for the associations and Cramér's V for their strength. Cohen's (1988) guidelines were used to evaluate the strength of the effects, with $0.1 < V < 0.3$ indicating a small, $0.3 < V < 0.5$ a medium, and $V > 0.5$ a large effect size. We report only significant effects, except where noted otherwise.

The **case of the pronoun** is associated with both the use of contraction ($\chi^2 = 230.02$, $p < 0.0001$, $V = 0.54$: large effect size) and the presence of emphasis markers ($\chi^2 = 12.47$, $p < 0.001$, $V = 0.13$: small effect size). This means that clefts are more contracted with accusative pronouns and less with nominative pronouns, and that emphasis markers co-occur significantly more often with nominative case pronouns. These effects are shown in Figure 3.

The **choice of complementiser** is associated with the case of the pronoun: nominative pronouns are used more frequently with *who* ($\chi^2 = 78.31$, $p < 0.0001$, $V = 0.31$: medium effect

[12] All $df = 1$, except when a Monte Carlo simulation is used (in this case, df are not applicable).

size). Moreover, the complementiser *who* co-occurs significantly more with uncontracted clefts ($\chi^2 = 19.46$, $p < 0.0001$, $V = 0.16$: small effect size). The two effects are visualised in Figure 4.

3.2.3.3 Syntactic role of clefted pronouns

Subsetting the data to the 148 clefts involving two referents, an obvious prevalence of clefts of the subject can be seen (112 vs 36, 75.7% vs 24.3%). While the subject is naturally a topical element in the information structure of an utterance, the object often provides new information, making it less necessary to mark explicitly with a focus.

In structural positions, the pronoun case is determined by the syntactic role the pronoun occupies (with nominative for subjects, and accusative for objects). However, this is not found in the clefted positions in our data. There is no significant association of the syntactic role of the clefted element with the case of the pronoun $\chi^2 = 3.67$, $p = 0.06$). As shown in Figure 5, both subjects and objects are more likely to appear in accusative case in our Twitter data.

No significant association of syntactic role with the use of contraction in the cleft was found $\chi^2 = 0.004$, $p = 0.95$), nor with the presence of emphasis markers $\chi^2 = 1.5$, $p = 0.22$). On the other hand, the syntactic role is associated with the complementiser: clefts of the subject co-occur significantly more with *who* rather than *that* ($\chi^2 = 15.47$, $p < 0.0001$, $V = 0.32$: medium effect size). This is visualised in Figure 6.

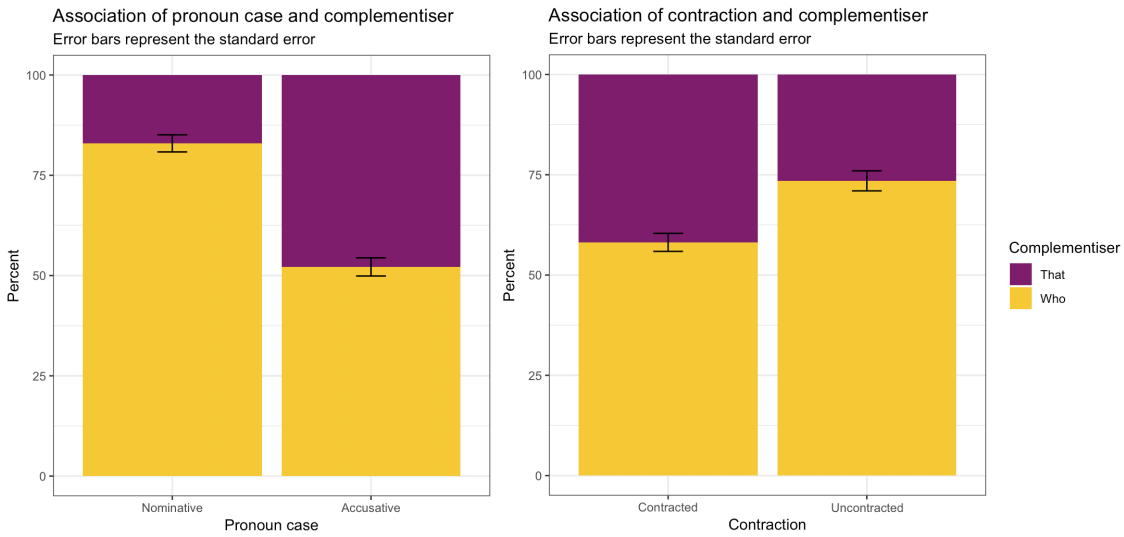

Figure 4. Associations of complementisers with case and contraction.

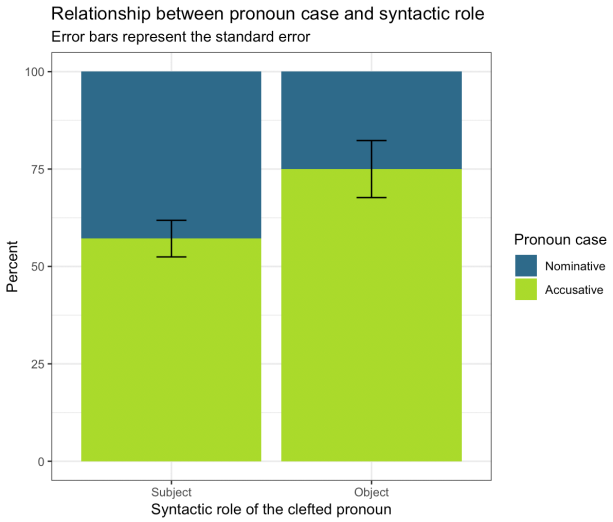

Figure 5. Distribution of case and syntactic role in clefts with two referents. The difference between subjects and objects is not statistically significant.

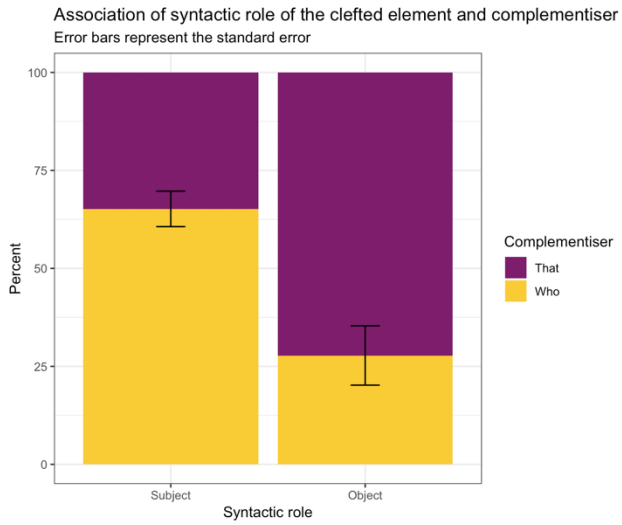

Figure 6. Associations of complementisers with the syntactic role of the clefted element.

3.2.3.4 Replication

The associations above were confirmed with a second, not manually annotated batch of 616 tweets mined on 8th May 2019. Table 1 compares percentages in the two datasets. Three of the chi-squared and Cramér's V tests could be replicated on the non-annotated data, yielding similar results (see Table 2). This replication shows that the results from our first analysis are stable across data samples.

Table 1. Associations of complementisers with the syntactic role of the clefted element.

	First batch	Second batch
Contracted | Uncontracted	60.8% | 39.2%	65.7% | 34.3%
Accusative | Nominative	61.0% | 39.0%	60.7% | 39.3%
That | Who	35.8% | 64.2%	42.6% | 58.4%

Table 2. Comparison of χ^2 and Cramér's V tests on the two batches of data. All $p < 0.001$.

	First batch	Second batch
Contraction ~ Case	$\chi^2 = 230.02,\ V = 0.54$	$\chi^2 = 261.99,\ V = 0.65$
Complementiser ~ Case	$\chi^2 = 78.31,\ V = 0.31$	$\chi^2 = 89.68,\ V = 0.38$
Contraction ~ Complementiser	$\chi^2 = 19.46,\ V = 0.16$	$\chi^2 = 27.95,\ V = 0.21$

Table 3. Twitter and iWeb comparison: model fixed effects. Each factor is reported by one of its two complementary variants (e.g. "Nominative" as one of the instances of case), and the estimates describe an increase (or decrease) on a logit scale.

Effect	Estimate	Std.Error	z value	Pr(>|z|)	
(Intercept)	1.09	0.29	3.75	< 0.001	***
Nominative	−0.74	0.30	−2.44	0.015	*
Uncontracted	−1.75	0.22	−7.79	< 0.001	***
Who	0.13	0.19	0.72	0.47	
Nominative:Uncontracted	−0.44	0.25	−1.73	0.08	
Uncontracted:Who	−0.26	0.28	−0.93	0.35	
Nominative:Who	−0.72	0.27	−2.61	0.009	**
	Significance thresholds: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$				

3.2.3.5 Comparison between Twitter and general web English

These associations were then recalculated on the results of corpus searches in the iWeb Corpus (Davies, 2018), a 14-billion-word corpus of written English data from 94k+ systematically-chosen websites.

Within the 4056 pronominal it-clefts found in the iWeb corpus, all associations seen in the Twitter data are replicated (with $p < 0.001$). Case is associated with contraction ($\chi^2 = 783.24$, $V = 0.44$) and with complementiser choice $\chi^2 = 362.28$, $V = 0.3$), and contraction is associated with complementiser choice $\chi^2 = 89.15$, $V = 0.15$). While the effects are more significant, likely because of the dimensions of the corpus, Cramér's V figures show slightly

smaller effects (especially for the association of contraction and pronoun case). Note that same direction of the effects will be confirmed through logistic regression.

The Twitter and iWeb results were compared through chi-squared tests in a 2 × 16 contingency table (including gender among the variations), computing p-values by Monte Carlo simulation (Hope, 1968) because of the low counts for some of the structures. A significant difference was found ($\chi^2 = 1827.5$, $p < 0.001$) between how Twitter and general-purpose (web) English use clefts of different types. Figure 7 shows how Twitter users adopt more accusative pronouns and contractions, while all-purpose language in iWeb attests at a more formal level avoiding contraction, as dictated by some style manuals, e.g. the Wikipedia Manual of Style (Wikipedia contributors, 2019). Nominative pronouns are used more outside of Twitter, mostly because of the very high frequency of the structure *It is he who* (forming 77% of all nominative-pronoun structures). Complementisers are used similarly across genres.

A generalised mixed-effects logistic regression (Bates et al., 2015) was fitted to predict the "genre" binary (Twitter vs iWeb) from the features of the clefts, i.e. case, contraction and complementiser, with a random effect for the gender of the pronoun (not assumed to display systematic variation across genres). All predictors were two-level factors. Model comparison was conducted via asymptotic likelihood ratio tests between a model without interactions, one adding the two-way interactions, and one additionally including the three-way interaction. Maximal random effect structure was used when supported by the data (as recommended by Barr et al., 2013): the random effects were successively simplified, chosen by lowest variance, until convergence was reached. The final model included a random intercept for pronoun gender. Adding the two-way interactions improved the fit ($p < 0.001$), while the three-way interaction did not ($p = 0.45$). The model results are reported in Table 3.

The model confirms that in Twitter clefts use fewer nominative pronouns ($p = 0.015$) and are more frequently contracted ($p < 0.001$). An interaction shows that clefts in which nominative pronouns and *who* co-occur appear less in Twitter than in iWeb ($p = 0.009$).

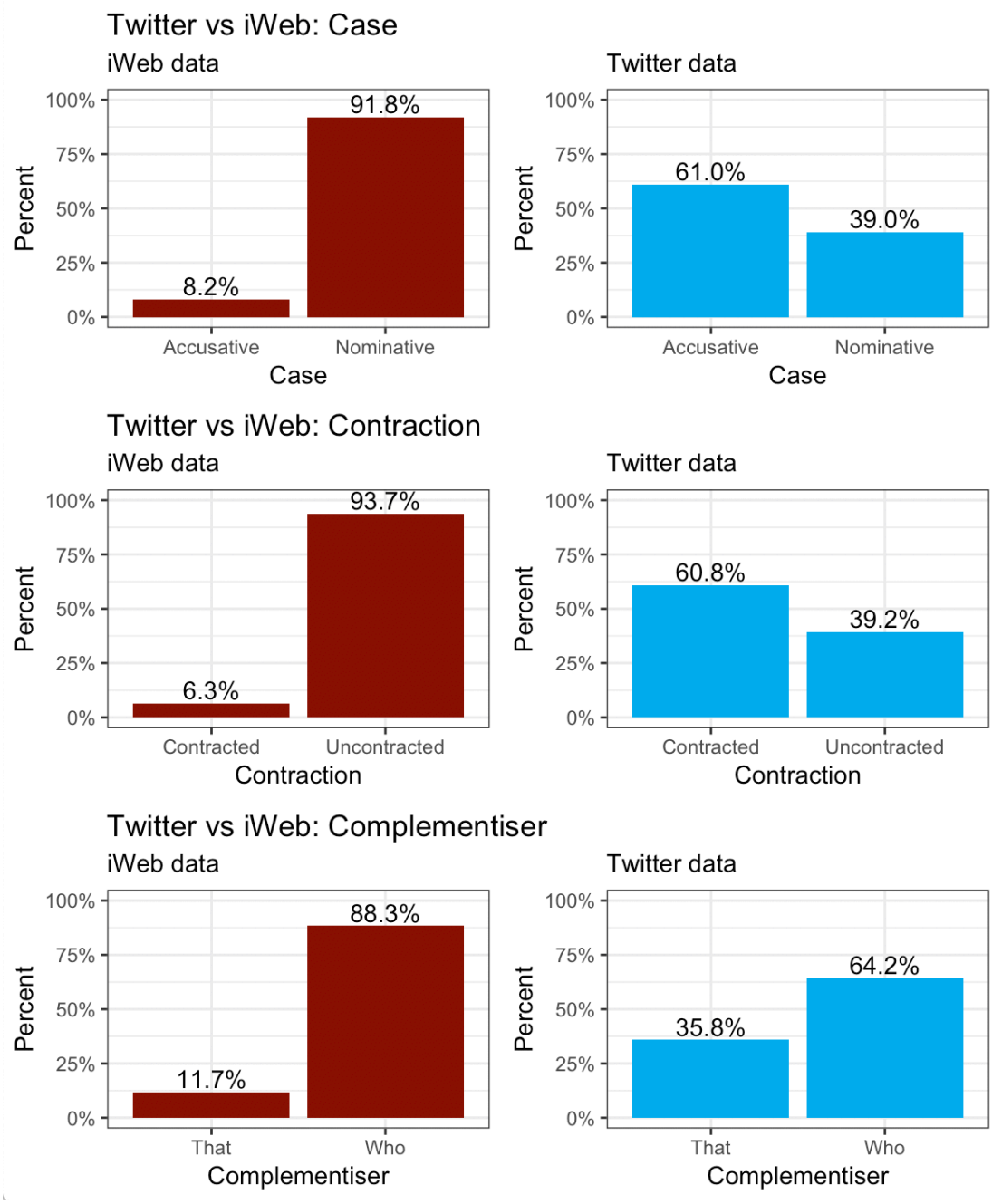

Figure 7. Comparison of iWeb with Twitter. The features on the right-hand side of each plot are more heavily prescribed in writing (nominative case, avoidance of contraction, and *who*).

3.2.4 Results

The results from the analysis in Section §3.2.3.2 confirm some of the hypotheses outlined in the introduction and in Section §3.2.2.3. The pattern that Cheshire et al. (2013) noted on relative clauses, whereby the complementiser *who* appeared most frequently with subjects, was replicated in it-clefts as the only association that the syntactic role of the clefted element showed: when an object was clefted, the complementiser *that* was used in almost three quarters of cases, while a clefted subject appeared significantly more with *who* as a complementiser (see Figure 6). While it is still not clear whether the case of the pronoun, and more specifically the nominative case, has a direct bearing on emphasis (as Maier 2014 suggests), an association was found between nominative pronouns and the presence of emphasis markers within the cleft (see Figure 3). This could mean either that two emphatic features are used concomitantly, or that the association is due to other features of the sentence that we are not able to see. Nominative case was moreover associated with uncontracted clefts, which rank high in the formality axis (while contraction has no effect on fuzziness and therefore does not directly influence emphasis). Our prediction in Section §3.2.2.3 that the complementiser *who* may be associated with the avoidance of contraction, stemming from the fact that this form ranked higher in formality and expressiveness, was confirmed (see Figure 4). This complementiser was likewise associated with nominative pronouns, which, however, cannot be explained as a less fuzzy version of accusative case pronouns. The hypothesis that the presence of emphasis markers could be associated with uncontracted clefts or *who* was not borne out. The difference outlined between Twitter and general-purpose internet English shows how social media language is more variable than edited writing. This is in keeping with the "conceptual orality" interpretation of social media language. The sources in the iWeb corpus show very stark tendencies, which may be due to the editorial process its texts undergo: most notably, the avoidance of contraction is almost ubiquitous (93.7%), and it is known to be one of the most prescribed-for traits of written language.

3.2.5 Discussion

Our main data on the variability of English it-clefts comes from written texts on the social media platform Twitter, which we compare to general purpose written English web texts.

We start from the situation that one of the variants is heavily preferred prescriptively in English writing, as confirmed by the edited web data. Contractions are typically avoided in written English (e.g. Wikipedia contributors, 2019: List of English Contractions).

Most of the current grammars of English discussing cleft sentences do not focus on the case of the pronoun, or on the distribution of the different form variants of the features we selected (e.g. Aarts, 2011; Hilpert, 2014; Cummings, 2018; Depraetere and Langford, 2019). Even though some mention that accusative forms are common in non-structural positions such as clefts (Cambridge Dictionary 2019; Eastwood, 1994: 62), most grammar advice in schools and online still prescribes nominative case for this construction (cf. Grammar Book, 2019; Quirk et al., 1985: 338 calls the accusative form "informal"). An exception to this are the Longman Grammar (Biber et al., 1999: 336), where both nominative and accusative forms are seen as possible but nominative pronouns are noted to be "presumably felt to be more correct" in formal registers, and the Cambridge Grammar (Huddleston and Pullum, 2008: 1414), which characterises the nominative as "formal or very formal" and the accusative as preferred in colloquial uses. Similarly, Maier (2013) shows that while nominatives dominate in written English corpora for clefts, spoken corpora exhibit many more accusatives.

As for the complementiser, most grammars mention both *who* or *that* for human referents and/or subjects, with *that* used in other cases (Eastwood 1994:62). The Longman Grammar (Biber et al. 1999) notes that in written registers, *who* is often chosen, while *that* or the omission of the complementiser are more typical of casual speech.

The prescriptively "correct", or at least preferable, version of the English it-cleft thus seems to be uncontracted, using a nominative pronoun, and (given that most clefted elements are human subjects), using the complementiser *who* (though *that* is also possible, in particular for objects).

Looking at formality and emphasis in more detail (as discussed in Sections §3.2.2.2–3.2.2.3), more formal texts generally avoid contractions, and, as noted above, tend to favour the nominative pronoun in the context of cleft. As for the complementiser, it is not

immediately clear which one is more formal. However, following the definition of formality in Heylighen and Dewaele (1999), more precise word forms are more formal than general-purpose words, because they avoid semantic fuzziness and context-dependence. Under this view, *who* is more formal than *that* when applied to human referents (as in all of our examples), since it is more precise in selecting specifically for humans.

The formal variants thus overlap with the prescriptively preferred variants. And indeed, the iWeb corpus of edited online written English shows the formal/prescriptive version of the it-cleft almost categorically, as shown in Figure 7: each formal feature is chosen in about 90% of the instances. This exemplifies a tendency that writers have, even in online texts, to adhere to formal style standards. The Twitter data is even more interesting in this light, since it shows considerable variation between the formal and informal variants.

As Heylighen and Dewaele (1999) note, expressivity/emphasis is somewhat orthogonal to (in)formality, and higher emphasis is associated with higher specificity but also higher context-dependence. Note that uncontracted clefts are able to receive more stress than contracted ones, and thus can be more emphatic. As argued above, *who* is more specific than *that*, thus also more emphatic under the assumed definition. Finally, Maier (2014) claims that nominative pronouns are more emphatic than accusative ones. This does not follow in Heylighen and Dewaele (1999)'s framework, as nominative pronouns are not more semantically specific or context-dependent than accusative ones. Here, we show that the case of the clefted pronoun in Twitter is not used to mark the syntactic role of the referent, which may free up the case feature to mark something else.

Though written, Twitter data tends towards a very informal style by default, exhibiting many features typical of conversational spoken language, and many informal variants (Eisenstein, 2013b; Hu et al., 2013; Koch and Oesterreicher, 1985; Storrer, 2013). In Twitter, using formal style features constitutes a marked case that, by being somewhat unusual, increases the general emphasis on the utterance. Taken together, we believe that the dimensions of (in)formality and emphasis at least in part explain the "clustered" distribution of cleft variants seen in Figure 2. First, we observe that the lack of enforcement of prescriptive standards leads to a much higher variability in this medium than other written (web)

text. Second, the formal variant that dominates in edited text (Uncontr-NOM-who) is still highly present in the Twitter data, but it is joined equally by two other variants. Of these, one reflects informal usage, which is default on Twitter (Contr-ACC-that). Further, the more specific complementiser *who* is also used often in informal Twitter clefts, increasing the emphasis in these cases (without seeming too formal, since both *who* and *that* are considered standard for clefts). Finally, further emphasis may be added by using the nominative pronoun and/or by adding orthographic emphasis markers, which are associated with nominative pronouns. Other combinations are rarer, possibly because of a resulting clash between the underlying utterance properties of (in)formality and emphasis.

In summary, in this paper we studied English it-clefts in written social media texts. We observe considerable variation in the realisation of the clefts with regard to the (non)contraction of the copula, the case of the pronoun, and the complementiser used. This variation is not random – instead we observe several close associations of the studied features, as well as the presence of additional orthographic emphasis markers. We conclude that while general web text is subject to overt standard language conventions, these are less relevant in Twitter and uncover different subtypes of it-clefts in form. We characterise these subtypes in terms of their placement on scales of formality and emphasis. It would be very interesting to see if the variation in form also corresponds to further functional distinctions among these it-clefts (for example, as far as their information structure or the relative salience of clefted referents in subsequent discourse). We must leave these questions for future research.

3.3 Conclusions

The current chapter started looking at prominence from a skin-deep point of view: before investigating how prominence works and affects language production or interpretation in the rest of this thesis, it examined the surface form that it takes in the shape of emphasis. The following chapters will progressively zoom in, looking first at the effect of prominence at sentence level (Chapter 4), then at event structure level (Chapter 5), and finally within the components of the meaning of an entity (Chapter 6).

Some of the study results can be related to theories surveyed in the literature review above (Chapter 2). Most notably, as predicted by Accessibility Theory (Ariel, 1990; see §2.3) and derivable from Grice's (1989) maxims (§2.1) and the Neo-Gricean accounts (§2.2), marked sentences correspond to marked expressions. The less common variant of a referring expression, that is the accusative form of the personal pronoun (*her* and *him*) was used along with emphasis markers. Forms that are either phonologically or semantically heavier (such as the uncontracted *it is* and the animacy-marked complementiser *who* rather than their respective counterparts *it's* and *that*) were also used to emphasise the focussed referent in cleft sentences.

However, these finding were found to vary with genre: in fact, when editorial rules intervene to normalise these variable features, the variants perceived as more formal are preponderant across the board. This was seen comparing clefts in Twitter and iWeb. In its greater variation (illustrated in Figure 2), Twitter language is closer to spoken language than edited standard English. It is possible that with it-clefts clearly marking emphasis, formal registers (unlike conceptually oral language) cannot support more nuanced variation in emphasis because of prescriptive rules. More effects of genre will be seen in Chapter 6: one of the studies, on English, compares the number agreement of named entities with anaphoric pronouns in production experiments and corpus texts at various degrees of formality and orality (from phone calls to web texts to newscasts). The results show that formality and the editorial standardisation of language are shown to influence the number agreement in the coreference of named entities, with formality correlating with the use of the singular pronoun *it* rather than *they*. The production experiments, moreover, score at an informal level, at an ambiguous position in the conceptual orality gradient akin to that of Twitter text.

In the next chapter (4), the same syntactic structure of clefted sentences is used to mark the prominence of a referent. Specifically, unmarked sentences are compared to clefts of the subject and of the object to see the effect of contrastive focus on coreference patterns, in interaction with the syntactic role of a referent. The role of information structure in next mention and referring expression interpretation and generation is variably seen as irrelevant (e.g. for Italian: Carminati, 2002; see also §2.5), as partially relevant (e.g. the Strong

form of the Bayesian model in Kehler et al., 2008; §2.5) or highly relevant, to the extent of being included in hierarchies formalising the accessibility of referents (e.g. Givón, 1983) and related to working memory (Almor, 1999; see also §2.4). Clefts provide the perfect case in point to examine the effect of prominence on coreference and referring expressions.

Chapter 4 The effects of prominence in coreference patterns

What are the repercussions of prominence marking?

4.1 Overview

While Chapter 3 was concerned with what prominence looks like at a surface level, investigating the ways in which it can be marked and whether different types of prominence marking stack up together in the same sentence or distribute complementarily in different genres, the current chapter will concentrate on what effects prominence has on the discourse, specifically on subsequent coreference patterns. In particular, two manipulations are used to affect the prominence level of referents: information structure, using the same focussing structure of the cleft sentence seen in Chapter 3, and sentence boundary. Much work in this area targets pronouns, and this study is a story continuation experiment in Italian, a language that gives participants the option to choose the most minimal referring expression of the null pronoun alongside overt pronouns and any other referring expression. Moreover, with a production experiment, participants are able to choose both who to talk about and what form to use to this end. This allows for the study of the effect of prominence along two different dimensions: choice of next mention and choice of referring expression, as well as exploring whether these two factors interact. Finally the effect of the prominence manipulations on coherence relations is examined as exploratory research.

The manipulations of information structure and sentence boundary make it possible to test the predictions of multiple theories delineated in Chapter 2. The presence of a sentence boundary updates the centres in Centering Theory (§2.3; Grosz et al., 1983), and the increased distance of the referent from its anaphoric expression makes the antecedent cognitively less activated and prominent (§2.4). In particular, referential distance is one of the main considerations in the Structure Building Framework (Gernsbacher, 1990), which predicts stronger referring expressions across longer distances; the Informational Load Hypothesis (Almor, 1999), similarly, talks about the "decaying" of referents over distance

using motivations related to working memory. All these theories would then predict the use of a heavier referring expression across sentence boundary.

Information structure is also an often-cited factor affecting coreference; however, topical referents are considered more often than focal referents, because stereotypically topical referents are at the same time subject, agent, and topic. Topic is one of the concerns of the Structure Building Framework (Gernsbacher, 1990), and information structure is (somewhat indirectly) considered in Accessibility Theory (§2.3; Ariel 1990). Focus is seen in divergent ways, both as something that increases a referent's prominence (with its prosodic prominence often cited as a factor; see e.g. Zimmermann and Onea, 2011) and as something that hinders it, being at polar opposites with topic (e.g. Givón, 1983). In other theories, such as the Position of Antecedent Strategy (§2.5; Carminati, 2002), information structure is explicitly given no role, as it is not a "structural" factor (that is, it is not grammatical in nature). Many of these theories, as outlined below (§4.2 and specifically §4.2.2), tend to merge interpretation and production, and with them the choices of next mention with that of referring expression, further mudding the predictions. Avoiding this, Bayesian models of coreference (§2.5; Kehler et al. 2008; Rohde and Kehler, 2014) predict information structure to only influence the choice of referring expression, and not that of next mention.

The following study on Italian, then, attempts to tease apart the effects that prominence (manipulated in the two ways described above) has on next mention and referring expression choice. A complex interaction of syntactic, information structural and other pragmatic factors is found in which it is impossible to establish either a mirror between the interpretive and productive patterns, or a clear division of labour between most minimal referring expressions such as null subjects and heavier ones such as overt pronouns. While the results confirm the expectations regarding sentence boundary, interestingly marking the prominence of a referent with a focus lead to two seemingly contrasting effects: while focussed referents were less likely to be mentioned again, the referring expressions used to refer to focussed antecedents were lighter.

4.2 Introduction

Studies on Italian referring expressions have mostly focussed on the null-subject nature of the language, which allows both null and overt pronouns along with other possibilities such as nouns and demonstratives. A frequently studied question has been how different features can influence the interpretation and perceived appropriateness and felicity of such pronouns. These theories have posited many factors to be the main reason for differing interpretations, sometimes from a mostly theoretical perspective, sometimes from an experimental perspective. Factors of a syntactic, semantic and pragmatic nature have been found: from grammatical role (mainly subject vs object, see e.g. Crawley et al., 1990) or, more formally, position in the derivation of a sentence (Carminati, 2002), to thematic roles (in particular agent and patient, or source and goal; e.g. Arnold, 2001), information structural or pragmatic notions (Huang, 2000; Levinson, 2000) and finally coherence-establishment processes (Kehler, 2002; Rohde, 2008; Kehler and Rohde, 2018). Less-formalised concepts such as prominence, salience or the influence of world knowledge have been taken into consideration, as well as biases such as the "role hierarchy preferences" or "grammatical role parallelism" (for a review of such theories, see Kehler, 2002).

Two assumptions that are often accepted to various degrees, or at least hinted at, are the division of labour between null and overt pronouns, and the mirroring between their interpretation and production. This double complementarity, however, may not hold across contexts and languages.

The clearest example of an account that postulates a division of labour between null and overt pronoun in Italian is that given by Carminati (2002) in her Position of Antecedent Strategy (PAS). In the PAS, the null pronoun preferably corefers with the most prominent antecedent, posited to be the referent in SpecIP position in Chomsky's (1981) Government and Binding syntactic framework, while the overt pronoun biases coreference to less prominent antecedents (those lower in the derivation). An example of the assumed null/overt usage is given in (1):

1) **Carminati (2002)'s predictions**:

   a. Maria$_1$ ha incontrato Sara$_2$ quando Ø$_1$ stava andando in biblioteca.
   Maria$_1$ has met Sara$_2$ when Ø$_1$ was going to the library.

   b. Maria$_1$ ha incontrato Sara$_2$ quando lei$_2$ stava andando in biblioteca.
   Maria$_1$ has met Sara$_2$ when she$_2$ was going to the library.

Most studies on coreference in null-subject languages, however, focus on interpretation, that is, on the comprehender's side. This conflates two different estimates: a) the hearer's estimate of *who* the speaker will mention, and b) the estimate of *how* the speaker will mention the chosen referent, or which referring expression will be formulated (for example, a null or an overt pronoun).

Studying cases like (1), one can see how, for comprehenders, an overt pronoun may signal that the target antecedent for pronoun resolution will not be the subject (1b). If the specific warning of the overt pronoun is not used, or in other words if a null pronoun appears instead, the hearer will go ahead and interpret the new clause as having the same subject as the previous one (1a). Still, these studies cannot give us a full picture in terms of the choice of who to mention versus that of which referential form to use.

To investigate these choices, it is necessary to look at the production side. Knowing what speakers may do in different contexts can in turn inform us about how comprehension works, as comprehenders will assume a typical speaker to behave coherently (though not necessarily mirroring the rules the comprehender follows).

The following sections review evidence regarding a division of labour between referring expressions in null-subject languages (§4.2.1), and on the information structural factors known to influence next mention and referring expressions (§4.2.2). The current study addresses the choice of antecedent and referring expressions in the production of Italian, attempting to discern which factors influence only the first, only the latter, or both, and further questioning how the choice of a specific form from a set of possible referring expression is made.

4.2.1 The null/overt division of labour

Carminati (2002)'s theory adopts the notion that the referents' relative prominence is influenced by syntactic position from Centering Theory (Grosz et al.,1983; 1995), but takes a further step, identifying the most prominent antecedent structurally. This means that in the PAS the antecedent bias is not influenced by non-structural factors (2002: pp. 57ff). Italian's null and overt pronoun then function in a division of labour whereby null pronouns correspond to the Spec IP position (the canonical subject) and overt pronouns correspond to other, less salient referents. The Spec IP is always the most prominent element, trumping the salience of positions that are higher, but more peripheral, in the syntactic derivation. These would include positions reserved for information structurally constrained NPs such as a marked topic or the focus positions in the CP domain. The PAS hence predicts that, say, a focussed subject, moving to the CP from the Spec IP position, will be easily retrieved by a null pronoun, while a focussed object, while also being in the CP, will need a more marked form like the overt pronoun in order to be referred to, because its movement did not start from the Spec IP (which is occupied by the subject).

Traditionally, this division of labour between null and overt pronoun stems from Chomsky (1981)'s Avoid Pronoun principle. Some corollaries of this principle applied to Italian state that when the overt pronoun is used, it bears contrastive properties (Rizzi, 1997) or is used to refer to an "unfamiliar" antecedent (Cardinaletti, 2004). Frascarelli (2007) further elaborates noting that strong (phonologically accented) overt pronouns can be used to cause a topic shift, while weak, unaccented overt pronouns correspond to nulls and, moreover, that a preverbal contrastive focus cannot be retrieved by a null pronoun. It can be noted that the Avoid Pronoun principle is characterised by Chomsky as "a subcase of the conversational principle of not saying more than is required" (1981: p. 65), thus bringing the discussion back at a Gricean discourse level, rather than a purely syntactic one.

Carminati (2002)'s findings are in fact in keeping with the frequent generalisation that the most salient or topical entity is referred to with the most reduced form of the inventory of referring forms. This association between the predictability of a referent and the reduction of the form used to refer to it has been tested experimentally. In a series of studies, Rosa and Arnold (2017) find that, in English, thematic roles influenced referential forms,

and specifically pronouns were used more to refer to goal entities (the recipients of the action), which are known to be the most predictable entity (see e.g. Arnold, 2011). Moreover, they report that goal entities were also considered more likely to be re-mentioned by the participants, concluding that predictability explains the effect of thematic role. Similar results are provided by Tily and Piantadosi (2009), who show that speakers are more likely to use reduced referring expressions when they are more certain about what entity will be mentioned next, and correspondingly that entities which are referred to with reduced forms are more predictable, even when participants are not able to see said form.

The model resulting from a clean-cut division of labour depicts a speaker and a comprehender who mirror each other: the speaker uses a reduced form to refer to the most prominent (or salient, or topical, etc.) entity *precisely because* the hearer will interpret this form to be coreferent with the most prominent entity. This mirror relationship has been the prevalent idea since at least the seventies (but see, even earlier, Grice, 1957), with claims that the production of pronouns can be simply modelled "by viewing production and perception as mirror-image processes" (MacKay and Fulkerson, 1979), and has been taken for granted in recent interpretation studies (e.g. Hemforth et al., 2010: "Exposure to these [referential forms] distributions will consequently shape preferences in comprehension"). However, these explanations collapse next mention choices and form choices, so the factors affecting each of them are not transparent.

A division of labour between the two types of pronouns in null-subject languages is implicit in most psycholinguistic experiments in pronoun resolution (e.g. Sorace and Filiaci, 2006), although its application to the interpretation of pronouns in null-subject languages has at times been found problematic. In Italian, Serratrice (2007) does not find a strong subject bias in the interpretation of the null pronoun: while monolingual adults interpreted the overt pronoun as coreferent to the previous object in 95% of cases, the null pronoun was linked to the subject in 46% of instances (vs 30% to the object). Likewise, in a study on the interpretation of pronouns in anaphoric and cataphoric contexts, Fedele and Kaiser (2014) point to the absence of a clear division of labour in subject pronoun resolution: an overall bias of the null pronoun towards subjects and of the overt pronoun towards non-subjects is confirmed; however, these are merely preferences and their

strength varies in different conditions, with the anaphoric condition strengthening the object bias of overts and the cataphoric condition strengthening the null pronoun's subject bias.

Similar discrepancies with the PAS have been found in studies on Spanish. A study by Alonso-Ovalle et al. (2002) on null vs. overt pronouns and the topic-focus articulation also shows an absence of division of labour: while a null pronoun was interpreted as coreferent with the subject in over 70% of cases, the overt pronoun was interpreted at chance level (50.2% of subject coreference); in a second study, both types of pronouns received bound variable interpretations (86% vs 63% of times) in sentences equivalent to the English "No student$_1$ believes that Ø$_1$/he$_1$ passed the exam". However, the authors claim nonetheless that Carminati (2002)'s PAS describes the behaviour of Spanish well basing their affirmation on the relative differences described above, as well as acceptability rating results. Their results are very similar to those of Filiaci (2010), a self-paced reading task using a Spanish translation of Carminati's materials, which confirms the null pronoun subject bias, but finds no strong connotation for the overt pronoun. Moving away from Romance null-subject languages, in a study on Japanese Ueno and Kehler (2016) find that both null and overt pronouns are subject biased.

Other evidence is pointing towards a separation between the factors influencing re-mention and those influencing form (Kehler et al, 2008), making a strict mirror model between interpretation and production untenable. Kehler and Rohde (Kehler and Rohde, 2018; Rohde and Kehler, 2014) characterise the relationship between pronoun interpretation and production in Bayesian terms:

2) $$P(referent \mid pronoun) \sim P(pronoun \mid referent)\ P(referent)$$

where the left term of equation (2) represents the interpretation bias that given a pronoun, it will be coreferent to a certain referent, while the right term corresponds to the likelihood P(pronoun|referent), the production bias that given a particular referent, a pronoun will be used to refer to it, multiplied by the prior probability that the referent will be mentioned at all. The model points to the fact that the prominence affecting interpretation and that affecting production need not be one and the same.

In a null-subject language like Italian, one must also consider other factors intervening in the economy of language versus precision of expression balance. In a perfectly Gricean mirror between production and interpretation, the speaker and the hearer share the same understanding of pronoun use and are both aware of the Maxims of Quantity and Manner (Grice, 1989), such that a richer expression will not be used for no reason because it may end up being misleading for the hearer, who assumes a reason for it to be used.

However, such Gricean-cooperative behaviour is not found in other aspects of language production. Speakers violate Quantity in not saying enough, preferring to lighten their cognitive load by mentioning available material as early as possible, rather than making further effort to prevent ambiguities (Ferreira and Dell, 2000). This would predict behaviour in which a dropped subject like the null pronoun in Italian would be the preferred referring expression even in cases when a richer expression would be able to disambiguate reference. Conversely, speakers have also been found to violate Quantity in the direction of over-informativity, an example being repeating a name or an overt pronoun even though this means incurring in a processing penalty (Pechmann, 1989; Gelormini Lezama, 2018): this would lead to a more indiscriminate use of heavy referring expressions.

Applying the idea that a less stereotypical referent will need a stronger referring expression than a more stereotypical one (with the term *stereotypical* henceforth used as an umbrella term for the prominent/salient/activated subject/topic), the production of referring expressions in Italian can be rewritten as (3):

3) $P(null\ pro|\ ref_{st}) > P(overt\ pro|\ ref_{st}) > P(proper\ name|\ ref_{st})$

where ref_{st} is the most stereotypical referent. Further intermediate steps between the minimal null and the maximally informative proper name would be possible along with the overt pronoun, with anaphoric expressions such as demonstratives or common nouns.

Another inference that can be made from a mirroring relationship between interpretation and production, also following from a principle of economy, would be that, since a lighter referring expression such as a null would be read preferably as coreferent with the most salient entity, the most salient entity would also be the most obvious next mention in the

discourse. To swap referent and continue the discourse talking about a referent that does not have maximal salience in the discourse, and hence is less palatable for re-mention, the speaker would use a non-minimal referring expression. In Italian, that would correspond to a non-null pronoun, as in (4):

4) $$Referring\ expressions = \begin{cases} null, & P(referent) = maximal \\ marked, & P(referent) < maximal \end{cases}$$

The distinction shown here represents the idea that salience has the same effect on next mention bias and referring expression: if a referent is very salient it will be re-mentioned with a null pronoun. This would be the production equivalent of the "topic shift trait" assigned to the overt pronoun by Sorace and Filiaci (2006) in their interpretation study, that is the property overt pronouns have to signal a change in discourse topic.

However, the Strong Hypothesis Bayesian model (Kehler and Rohde, 2018) predicts such a linear use of referring expressions not to occur. In the system above (3–4), there is a perfect trade-off wherein maximal salience corresponds both to maximal re-mention probability and to minimal referring expression: the same factors that influence salience of mention also influence referential form.

In contrast, the Strong Hypothesis shows that some pragmatic factors, like information structure, bias the production of referring expressions but have no bearing on choice of next mention (Rohde and Kehler, 2014). Rather, the next-mention bias would be influenced by factors such as verb semantics and coherence relations (Kehler, 2002). The process of coherence establishment alters the salience of referents. On the other hand, referential salience has sometimes been shown to influence the referring expressions used to refer to them, with a non-Bayesian approach (as seen in Rosa and Arnold, 2017; but see Fukumura and van Gompel, 2010, for counterevidence), implying that in turn there would be an impact of the factors that influence salience of next-mention also on referring expression generation. This difference in the scope of action of different factors can best be

verified by manipulating one of the factors, such as the information structural status of a referent as signalled by syntactic constructions, while keeping the other constant.

This Bayesian approach has been primarily developed in English (but see Mayol, 2018, for an application to Catalan), and to its inventory which only allows for overt subject pronouns (typically signalling topic continuity). An open question, however, remains about how Kehler et al.'s (2008) observed separation of factors carries over to contexts like Italian, where the pronoun inventory is different.

4.2.2 The influence of information structure

The salience of a focussed element can be split in two and seen from the perspective of the speaker and that of the hearer. Unlike a discourse topic, which, being what the discourse is about, is shared by both parties of a conversation, a focussed referent is highly salient for the speaker (who will in fact choose to mark it in some way), but not salient, or even not known up to that point, to the hearer (cf. Lambrecht, 1994; Vallduvi, 1993). This asymmetry could in fact lead to a different performance in coreference resolution from the hearer and the speaker.

On the one hand, the focussed element is naturally emphasised, which could make it both favoured for re-mention and more likely to be referred to by a minimal form. Zimmermann and Onea (2011) predict that a focussed element should be more likely to be taken up and incorporated into the common ground: this would mean that, for example, an element that's contrasted as the focus of a cleft construction is part of a set of entities that is cognitively more relevant to the processing of a sentence – and, in turn, this would make the referent relatively more prominent. These results were found in an eye-tracking study in English by Kaiser (2011), in which contrastive focus boosted the salience of a referent even when a competing very salient referent (a pronominalised discourse topic) was present. Nonetheless, in other studies on anaphoric relations, such as one on English pronoun resolution between sentences by Cowles et al. (2007), both marked focus and topic are found to increase the prominence of a referent, which will in turn be apt to be retrieved by a pronoun; likewise, in their online study all marked targets (both by topic and focus) were named faster than unmarked referents. These studies, however, do not distinguish next

mention biases from choice of referential form, rather collapsing them into a single coreference category.

On the other hand, if the typical discourse continues to be about a topical referent, focussing an element could have the effect of making it less accessible, and thus disfavoured for coreference. Focus was in fact shown to make an antecedent less retrievable in a study on intrasentential anaphoric pronouns in Spanish (de la Fuente and Hemforth, 2013). The prominence-enhancing effect of contrastive focus seems thus to only take effect between utterances: focus strengthens the prominence of an element, so that when the common ground is updated for a new sentence it is easier to ratify it and transform it into the topic in the next sentence. Similar results are reported in questionnaire studies conducted in French and German (Colonna et al., 2012; 2010), in which subjects focussed through clefting showed reduced accessibility within a sentence.

Other than in interpretation studies, information structure has also been shown to have an effect on the production of referring expressions in English, with pronouns being the preferred production for referents made topical by being the subject of a passive-voice construction (Rohde and Kehler, 2014).

Notice that the contradictory results of the aforementioned interpretation studies may point to an interaction between the effects of focus and continuation type: while focus enhances the accessibility of a referent between sentences, with the prominent referent being picked up as the next topic (or centre in Centering terms), it reduces it within the sentence. Moreover, studies show that simpler salience factors can describe the behaviour of referring expressions cross-linguistically, relegating cross-linguistic variation to language-specific features in a pronominal system (for an extensive review, see Huang, 2000). If a mirror holds between the interpretation and production of pronominal forms, it would be possible to see such an interaction in a production study. The controversial status of focus as either prominence-enhancing or non-prominent may further reflect a collapse of two choices: that of which antecedent to re-mention and that of which referring expression to use to refer to it.

Problematising the matter further, self-paced reading results show that in Spanish the penalty found when participants encountered overt pronouns and names in infelicitous contexts between sentences can be eliminated when the target reference is emphasized through a cleft structure (Gelormini Lezama and Arnold, 2011). The same penalty was replicated in Italian (de Carvalho Maia et al., 2017) and was found to occur both between and within sentences, confirming both an overt pronoun penalty and a (weaker) repeated name penalty in the processing of anaphora (Gordon et al., 1993). The study also claimed that the weaker penalty found for names could be due to the fact that while names can foreground or emphasize information when a referent is salient, pronouns do not have any of these specific functions and their use is thus less justified.

The interplay between information structure (and specifically focus) and sentence boundary needs further study, both to clarify the interaction of these two factors in influencing anaphoric patterns and to confirm or disconfirm mirroring in the behaviour of pronoun interpretation and production.

4.3 Story continuation study

4.3.1 Motivations and predictions

The following study manipulates information structure and continuation type to test their impact on next mention bias and on the production of referring expressions. The study tests different models predicting referential form (the first) or its relationship with next mention biases (the latter two):

- **Position of Antecedent Strategy**: subjects are the most prominent element, regardless of information structural factors; they will thus be re-mentioned with null pronouns, the most reduced form in Italian's inventory (Carminati, 2002; see also §2.5);
- **Predictability ~ reduction models**: the more predictable a referent is, the more likely it is to be re-mentioned and referred to with a reduced form (e.g. Rosa and Arnold, 2017, and Ariel, 1990; see also §2.3 for an overview);
- **Strong Bayesian model**: biases for next mention will be distinct from biases for the choice of referring expressions, and only the latter will be influenced by

the information structure of the preceding clause (Rohde and Kehler, 2014; see also §2.5).

In the current study, an unmarked baseline is compared to two clefting structures, one focussing the subject and one focussing the object of the main clause. The continuation type conditions include a new sentence, to test for between-utterance effects, and a temporal subclause with *quando* ("when"), for within-utterance effects of information structure. While continuation type is not predicted to have an effect on next mention choice per se, its interaction with information structural modifications can provide insight into whether the effects observed in studies on pronoun interpretation and focus can be translated to pronoun production. If a lack of an effect of information structure on next mention were to be observed along with a relevance to referring expression choice, the results could furthermore give an insight into whether Kehler and Rohde's (2018) Strong Hypothesis, which postulates this difference in the factors affecting each, applies to Italian.

To give an insight on the full range of referring expressions, continuations were free, such that participants could continue the items with whichever referring expression they chose. Anaphoric expressions that have been somewhat neglected in experimental studies, such as proper names, demonstrative pronouns or common nouns used anaphorically (referred to here as *strong expressions*), were of particular interest. While the effect of continuation type was expected to be straightforwardly predicted by multiple theories (like Centering and Accessibility Theory, see §2.3), with stronger expressions used between-sentence and null subjects used preferably within-sentence, the effects of focussing a referent were the main object of interest: looking at both next mention and referring expression, the role of focus in changing the salience of a referent can be discerned from that of syntactic role.

With regard to next mention, while in syntactic role subjects are more likely to be mentioned next as a default topic continuation, marking either the subject or the object with a contrastive focus disrupts this linearity. Considering referring expressions, neither Carminati (2002)'s PAS nor "predictability ~ reduction" models predict the use of null pro-

nouns to refer to a focussed object. In both cases, a heavier referring expression was predicted, because the referent was respectively not in SpecIP position (occupied by the non-focussed subject) nor predictable (being neither a subject nor a topic). The Strong Bayesian model, on the other hand, predicts the influence of information structure on the choice of referring expression.

If a different effect of focus on next mention bias and referring expression is found, this could point to the two dependent variables not being as intrinsically linked as previously thought in models of Italian. While a referent that is treated as the topic may be favoured for re-mention and to be preferentially referred to with a minimal form (see e.g. for French and German: Colonna et al., 2012; for Italian: Carminati, 2002), focus could have a more nuanced and multidirectional effect: specifically, it may only increase next mentions across sentences but not within sentences (where the common ground has not undergone an update: Zimmermann and Onea, 2011) – this can also be expect if a comparison of the results of de la Fuente and Hemforth (2013) and Cowles et al. (2007) is made despite the difference in language: their results go in opposite directions with regard to the effect of focus on next mention but their design considers respectively within and between sentence anaphora, which could be the main source of the difference. As for the choice of referring expression, the predictions are more ambivalent (see §4.2.2).

Finally, the effect of continuation type and information structure on participants' choice of coherence relations is also analysed. An effect of continuation type is expected on the frequency of certain relations (such as Occasion) because of the intrinsic semantic meaning of *quando*: it is easier to express a series of events following in time with "when" rather than a cause-effect relation, which would prefer connectors equivalent to "because" or "thus". The effect of information structure on coherence relation has, to our knowledge, not been tested psycholinguistically. If information structure, keeping meaning (verb) equal, affects the chances of re-mention of a referent, the coherence of the text may be affected as well: rather than selecting a coherence relation that is biased towards a certain (subject or object) antecedent, the speaker could choose a relation that is most suitable to the referent elected to subject of the continuation.

4.3.2 Materials

The 36 experimental items consisted of a three-sentence paragraph each. These included an introductory sentence in which two referents of different gender A and B are mentioned in a coordinated NP, a second sentence which fills in more context concerning the referents, and a target sentence followed by a gap to be completed. All target sentences included A (subject), a transitive verb, and B (direct object). They all used the *passato prossimo* tense (akin to the English present perfect), and an adverbial or a prepositional phrase either preceded or followed the two referents. Proper names for the test referents were taken from the most popular male and female names for newborns in Italy in 2000 (Istat, 2018).

The experiment was set up in a 2x3 design, in which the target sentences to be completed varied two prompt continuation types and three information structure types. The information structure could be a non-altered baseline (5a), or a cleft structure focussing the subject (5b) or the object (5c). The continuation could either be a temporal subclause introduced by *quando* ("when"), or a new sentence. (6) gives an example of an item, complete with the two scene-setting sentences.

5) **Conditions and item examples**:
    a. BASELINE CONDITION:
       Matteo ha cercato Veronica al bar, quando...
       Matteo looked for Veronica at the bar, when...
    b. SUBJECT FOCUS CONDITION:
       È Matteo che ha cercato Veronica al bar, quando...
       It's Matteo who looked for Veronica at the bar, when...
    c. OBJECT FOCUS CONDITION:
       È Matteo che Veronica ha cercato al bar, quando...
       It's Matteo who Veronica looked for at the bar, when...
6) Matteo e Veronica erano andati in vacanza, al mare, dopo molti mesi di lavoro ininterrotto. Non avevano potuto prendere ferie per Natale, e il resort era un paradiso. Un pomeriggio, Matteo ha cercato Veronica al bar, quando...

Matteo and Veronica had gone on holiday to the seaside, after many months of uninterrupted work. They hadn't had a chance to take a break for Christmas, and the resort felt like a paradise. One afternoon, Matteo looked for Veronica at the bar, when...

The experimental items were interleaved with 36 fillers. These also included at least one human referent.

4.3.3 Participants and procedures

A total of 139 Italian native speakers aged 19-65 (mean age 28, σ = 6; 98 female) were recruited from Italian universities and received 8€ for an estimated 45 minute-task. Some of the participants had to be excluded insofar as they answered with nonsense or copy-pasted continuations, bringing the number of participants whose data could be analysed to 120. Not all participants completed the study in its entirety, but data from all the participants was included.[13] A pilot study was conducted on 5 participants prior to the experiment.

Continuations were collected via a web-based interface that participants could access from their own computer. The website displayed a background questionnaire, a consent form and an instructions page, and then proceeded to display one item at a time with a text box for participants to write their continuations in.

4.3.4 Annotation

Continuations for experimental items were annotated for the referent that participants used as the subject in their new sentence, for the referring expression used, and for the coherence relation expressed by the continuation.

Some of the referring expressions were excluded from all further analysis because of their scarcity in the data: the "nominative" case personal pronoun *egli/ella* (one occurrence of each), the demonstrative pronouns *questo/questa* (occurring respectively twice and once), and various common nouns (14 total occurrences). This left a total of 1852 continuations

[13] The mixed-effects models used for the analysis (as described in §4.3.5) are not affected by the different number of data points provided by each level of the participants factor. Moreover, they are also assumed to be robust to violations of distributional assumptions (Schielzeth et al., 2020).

for analysis: 1020 null pronouns, 501 "accusative" case personal pronouns (of which 271 were the feminine form *lei* and 230 the masculine *lui*), and 331 proper names.

Note that Italian has undergone a change whereby accusative third person pronouns have extended their use to the nominative function, making personal pronouns in the nominative virtually disappear (as shown in the continuation data). The use of the nominative forms *egli* and *ella* is still actively prescribed in education, but is not reflected in real language use since at least the XIX century (Berruto, 2017; Renzi, 1994, 2012).

Coreference was annotated by the author. Coherence relations, on the other hand, were labelled by the author plus an additional annotator (also a native Italian speaker and a linguist) because of their less clean-cut nature. Disagreements in the annotation were resolved when possible: in all other cases double annotations were kept, but not analysed further given their low frequency. The set of relations to annotate included six relations: five from the inventory described in Kehler (2002), plus a more general Background relation for information that was somehow related to the event of the preceding statement but was not described by any of the other relations. The full set of seven coherence relations includes:

- **Result**: cause-effect relation. The event follows as a result of the one in the preceding statement, or is implied by it.
  e.g. *John pushed Mark. Mark fell to the ground.*
- **Explanation**: cause-effect relation. The event is the cause of the one in the preceding statement, or implies it. Reversal of the Result relation.
  e.g. *Mark fell to the ground. John had pushed him.*
- **Violated Expectation**: cause-effect relation. The actual event is contrasted to another that would have been normally expected as a result of the preceding statement. A Result relation in which the result is unstereotypical.
  e.g. *John pushed Mark. Mark thanked him.*
- **Occasion**: contiguity relation. The event follows in time that of the preceding statement, but is not implied by it (like e.g. Result).
  e.g. *John stood up. He started singing.*

- **Elaboration**: resemblance relation. A restatement giving further information or detail on parts of the same event.
  e.g. *John started singing. He practiced scales.*
- **Parallel**: resemblance relation. Two referents carrying out two parallel arguments, that is, two events which are subsumed by a common property.
  e.g. *John started singing. Mark started playing the piano.*
- **Background**: a general relation for background information that is related to the event of the preceding statement in a non-specific way.
  e.g. *John started playing the piano. He had been studying it for many years.*

4.3.5 Results

The statistical analysis used generalised mixed-effects logistic regressions (Bates et al., 2015). The details of each model are described in their relevant section. The maximal random effects structure was used when supported by the data (Barr et al., 2013) and when a principal components analysis of the random effect did not indicate any over-specification. Where a model did not converge, the random effects were successively removed, chosen by lowest variance.

4.3.5.1 Analysis of next mention biases

The following analysis (including that of referring expressions) reduces the data set to the 1487 continuations which can be clearly attributed to either referent of the prompt. This means that the 365 continuations in which the null pronoun presented an ambiguous coreference were excluded (representing 36% of all valid null pronoun continuations).

The selection of the referent chosen to be the subject of the continuation was modelled using a generalised mixed-effects logistic regression (Bates et al., 2015) with the syntactic status of the referent (isSubject) as the outcome variable and fixed effects of Information Structure (infoStr) and Continuation Type (contType) of the prompt. All factors were centred. Because infoStr is a three-way factor, it was centred as the difference between pairs of its values, i.e. "focObj – baseline" and "focSub – focObj".

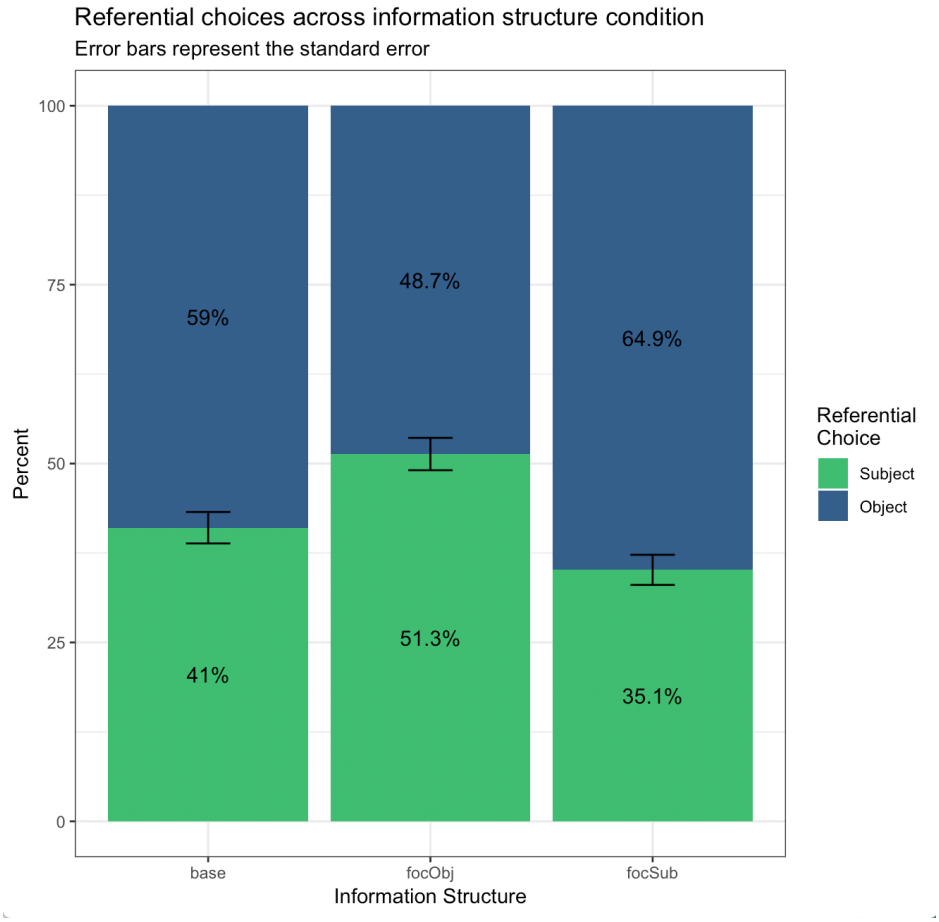

Figure 8. Antecedent re-mention across information structural conditions.

The model with Information Structure (baseline vs subject focus vs object focus) and Continuation Type (*quando* vs new sentence) as predictors was compared with one also including their interaction through a significant likelihood ratio test as well as a comparison of AIC and BIC. While the likelihood ratio test weakly preferred the model with the interaction, the AIC and BIC strongly preferred the one without interaction, showing the model with the interaction of the two predictors is not worth the additional model complexity ($p = 0.76$). The model without interaction was thus chosen.

The maximal converging model included random slopes for Continuation Type and its interaction by participant and prompt.

The data for next-mention biases is visualised in Figure 8. As can be seen, there is a bias against re-mentioning the subject and one against re-mentioning a focussed referent: the highest rate of subject re-mention is found when the object is focussed (51.3%), and the

lowest rate is found when the subject is focussed (35.1%), with the baseline attesting in between (41%). Overall, thus, the data shows a re-mention bias for non-focussed referents. Note that the Continuation Type predictor is not plotted because it did not show any significance in the modelling phase. The model's outcome is reported in Table 4.

Table 4. Subject re-mention: Estimated model fixed effects.

Effect	Estimate	Std.Error	z value	$Pr(>\|z\|)$	
(Intercept)	−0.26	0.13	−2.06	0.04	*
focObj - baseline	0.45	0.14	3.21	0.001	**
focSub - focObj	−0.67	0.14	−4.8	< 0.001	***
Continuation type	−0.17	0.16	−1.02	0.31	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

The model shows an overall bias towards mentioning the object of the prompt sentence (significant intercept, p = 0.04). While Continuation Type does not show any significant effect on antecedent selection (p = 0.31), information structure shows a difference between the three conditions: the selection of the subject antecedent increases focussing the object of the prompt (p = 0.001), and the two focussing conditions show the most difference in this respect (p < 0.001). This points to a next mention bias in which focussing an antecedent decreases its chance of re-mention, showing an opposite direction to the effect of topicalising a referent as shown in the literature (e.g. for French and German: Colonna et al., 2012; for Spanish: de la Fuente and Hemforth, 2013).

The effects of Information Structure were confirmed by subsetting the data into pairs of two of the three possible information structural conditions and fitting new generalised mixed-effect models (fitted without a random effect for participant, as this prevented the models from converging; a PCA confirmed that the random effect was not necessary):

- in the subset with the baseline and focussed subject conditions, Information Structure does not have a significant effect on referential choice (p = 0.128);
- in the subset with baseline and focussed object conditions only, the effect of Information Structure reaches significance (p = 0.002);

- in the subset with the two focussing structures, Information Structure shows a very significant effect ($p < 0.001$).

Two (simplified) continuations from the data, exemplifying the information structural effect, are given in (7) (object preference in the baseline condition) and (8) (subject preference in the focussed object condition):

7) Maria ha abbracciato Gabriele. Lui colse la palla al balzo e la baciò.
   Maria hugged Gabriele. He took his chances and kissed her.
8) È Gabriele che Maria ha abbracciato. Era impressionata dal film.
   It's Gabriele that Maria hugged. She was shaken by the film.

Summarising, these results show a strengthened re-mention bias for non-focussed referents. This is in line with models in which a focussed referent is not likely to be re-mentioned because it is a new introduction to the discourse and, as such, not a ratified, predictable element. Surprisingly, an interaction of information structure and sentence boundary did not describe the data more closely: this rejects the prediction that a focussed referent would be picked as a next mention more across sentences than within sentence, after an update of the common ground. The results, at the same time, are in contrast with the Strong Bayesian model (Rohde and Kehler, 2014; see also §2.5), in which information structure is predicted to influence the choice of referential expression rather than the choice of referent. As Carminati (2002)'s PAS does not discern between re-mention bias and referring expression bias, this section of the study cannot address its predictions.

4.3.5.2 Analysis of referring expressions

A summary of the distribution of referring expression across syntactic and focus status of the chosen referent is given in Table 5 and visualised in Figure 9. As shown in Figure 9, object coreference biases participants towards the choice of non-null referring expressions. The heaviest referring expression, that is the repetition of the proper name, is used more in new sentences than within sentence. Finally, when the subject antecedent is chosen, this is referred to with heavier expressions if the object is focussed, and focussing the sub-

ject too seems to slightly increase the rates of non-null expressions, especially across sentences. On the contrary, when the object antecedent is selected its anaphoric expressions tend to be lighter when the object antecedent was focussed than when it was not.

To model the effect of Information Structure (baseline vs subject focus vs object focus), Continuation Type ("when" vs new sentence), referent grammatical role (subject vs object), and referent focus (focus vs non-focus) on the choice of referential expression, three generalized linear mixed-effects models were defined following a step-down procedure analogous to the one used analysing antecedent choice (§4.3.5.1). Starting from a maximal model with all the interactions and maximal random effect structure, the three chosen models were:

- **isNull and isOvert models**: Fixed effects for Information Structure, Continuation Type, the referent's grammatical role and focus status, and the interaction of Information Structure with grammatical role, Continuation Type with grammatical role and of Continuation Type with focus, with random slope for participant and prompt and their interaction with Continuation Type;
- **isName model**: Fixed effect for Continuation Type, referent grammatical role and focus, with random slope for participant and its interaction with Continuation Type and a random slope for prompt.

Bonferroni's correction was applied to the p-values to allow for the fact that the dependent variable of referring expression was modelled three times. This means that the lowest p-value limit considered is 0.0003 rather than 0.001. The Bonferroni-corrected significance levels are reported in Table 6, along with the models' outcome.

Table 5. Distributions of referring expressions across conditions (percentages add up to 100% by line)

			Null	Overt	Name
Subject	When	Baseline	92.7	5.2	2.1
		Foc Sub	92.1	6.7	1.1
		Foc Obj	68.0	28.7	3.3
	New sentence	Baseline	84.6	1.8	13.6
		Foc Sub	63.0	14.1	22.8
		Foc Obj	42.4	22.4	35.2
Object	When	Baseline	11.6	73.6	14.8
		Foc Sub	18.2	62.4	19.4
		Foc Obj	54.8	34.1	11.1
	New sentence	Baseline	16.7	37.7	45.7
		Foc Sub	14.1	41.7	44.2
		Foc Obj	32.1	30.2	37.7
			44.0	33.7	22.3

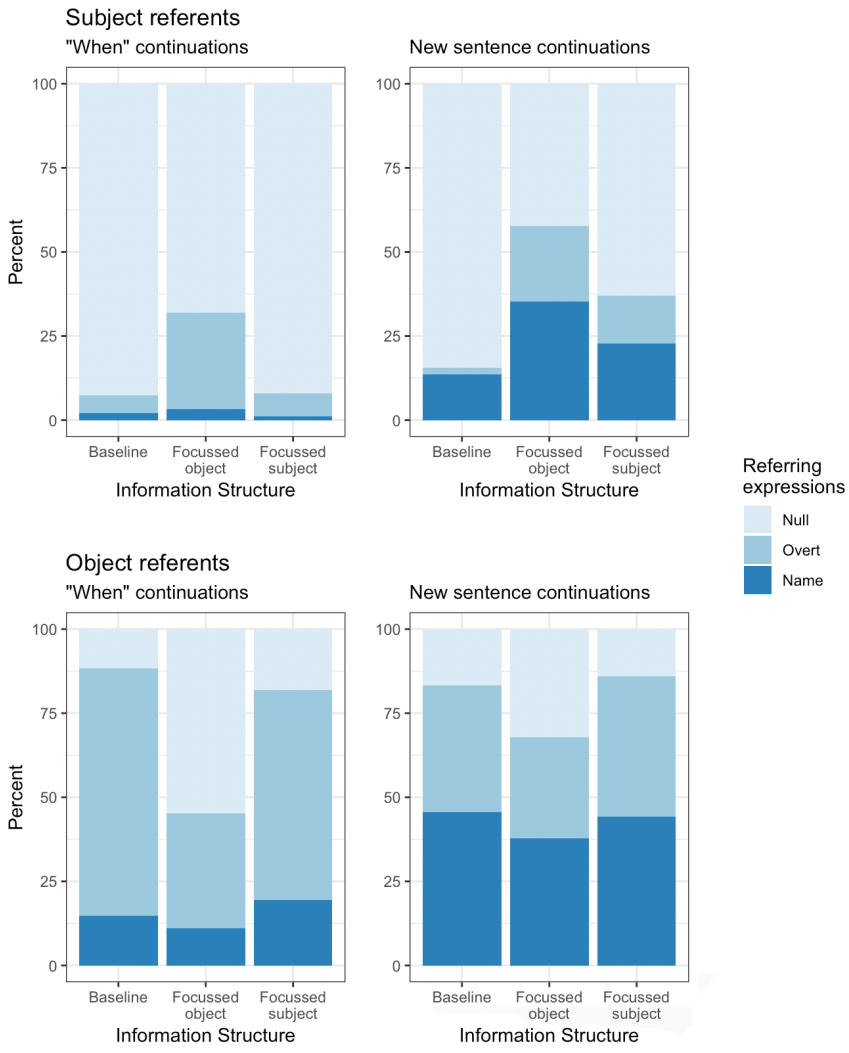

Figure 9. Distributions of referring expressions across conditions.

Table 6. Choice of referring expression: Estimated models' fixed effects.

Effect	Estimate	Std.Error	z value	$Pr(>\|z\|)$	
			isNull		
(Intercept)	0.19	0.13	1.43	0.15	
focObj - base	−0.98	0.27	−3.65	< 0.0003	***
focSub - focObj	−0.14	0.24	−0.62	0.55	
contType	1.3	0.26	4.9	< 0.0003	***
isSubject	3.28	0.24	13.83	< 0.0003	***
isFocus	1.6	0.22	7.4	< 0.0003	***
focObj - base : isSub	−2.89	0.44	−6.6	< 0.0003	***
contType : isSub	0.62	0.44	1.4	0.16	
contType : isFocus	0.91	0.4	2.26	0.02	
			isOvert		
(Intercept)	−1.79	0.19	−9.3	< 0.0003	***
focObj - base	1.15	0.3	3.89	< 0.0003	***
focSub - focObj	−0.06	0.23	−0.26	0.8	
contType	0.49	0.27	1.8	0.07	
isSubject	−2.68	0.23	−11.65	< 0.0003	***
isFocus	−1.19	0.22	−5.42	< 0.0003	***
focObj - base : isSub	2.72	0.5	5.44	< 0.0003	***
contType : isSub	−1.08	0.41	−2.66	0.008	*
contType : isFocus	−1.41	0.39	−3.62	< 0.0003	***
			isName		
(Intercept)	−2.61	0.26	−10.1	< 0.0003	***
contType	−2.51	0.36	−7.04	< 0.0003	***
isSubject	−1.58	0.21	−7.41	< 0.0003	***
isFocus	−0.54	0.22	−2.52	0.01	*

Significance with Bonferroni correction: $^{***}p < 0.0003$, $^{**}p < 0.003$, $^{*}p < 0.017$

Starting from the independent variables in the study (visualised in Figure 9), Continuation Type had a significant effect on null pronouns and names (both p < 0.0003), whence null pronouns occur more within sentences (*quando* continuations) and names occur more between sentences (new sentence continuations). No significant effect of Continuation Type was found on the use of overt pronouns, such that most of the trade-off seems to take place between dropping the subject (preferred referring expression in the "when" continuations) and expressing it with a proper name (increased use in the new sentence continuations). The use of heavier referring expressions between sentences is predicted by "predictability ~ reduction" models (§2.3).

Information Structure showed a significant effect on the use of null and overt pronouns, while its influence was not modelled in the isName model, since it did not significantly improve its fit (p = 0.06). The significant effects reflect a difference between baseline and focussed object conditions, such that focussed object prompt yields more null pronouns

and fewer overt pronouns than the baseline (both $p < 0.0003$). The influence of information structure on referring expression selection is predicted both by "predictability ~ reduction" theories (although in different ways for different theories, based on whether a focus is considered an enhancer or deterrent for prominence: see §4.2.2) and by the Strong Hypothesis of the Bayesian model (Rohde and Kehler, 2014).

These effects of Information Structure on null and overt pronouns were confirmed by subsetting the data into pairs of two of the three possible information structural conditions and fitting new generalised mixed-effects models. The results show that the greatest difference is to be found between the baseline and the focussed object condition:

- **Null pronoun sensitivity to Information Structure**: to follow up the significance of Information Structure in the full model (see Table 6), each pair of Information Structure conditions was compared. The only reliable difference was the higher rate of null pronouns in the Baseline condition over the Object Focus condition ($p < 0.001$) (as compared to Baseline vs SubjectFocus: $p = 0.464$; and SubjectFocus vs ObjectFocus: $p = 0.71$);
- **Overt pronoun sensitivity to Information Structure**: with regard to the overt pronoun, the only reliable difference was the lower rate of overt pronouns in the Baseline condition over the Object Focus condition ($p < 0.001$) (as compared to Baseline vs SubjectFocus: $p = 0.221$; and SubjectFocus vs ObjectFocus: $p = 0.752$).

If the chosen referent is focussed or a subject, the same effects are achieved on referring expressions: the occurrence of the null increases (both $p < 0.0003$), while the occurrence of overt pronouns ($p < 0.0003$) and names (isSubject: $p < 0.0003$; isFocus: $p = 0.01$) decreases (see Figure 10).

Some interactions also reach significance. In particular, the interaction of isSubject:infoStr inverts the general tendency – with a subject antecedent, focObj yields fewer null and more overt pronouns than the baseline (both $p < 0.0003$). This effect is visualised in Figure 11.

Continuation Type, on the other hand, interacts with both isSubject and isFocus enhancing the previously mentioned main effects on overt pronouns: with a when, within-sentence continuation, subjects and focussed referents yield proportionally even fewer overt pronouns ($p = 0.008$ and $p < 0.0003$ respectively). These interactions are visualised in Figure 12.

The influence of information structure on referring expressions is predicted by the Strong Bayesian model (Rohde and Kehler, 2014). Conversely, the results go counter the predictions of Carminati (2002) and of "predictability ~ reduction" models, which do not expect a null pronoun to corefer to a focussed object, as this antecedent is neither in the SpecIP position (i.e. it is not a subject), nor is it a topic.

Moreover, the comparison between the choice of next mention (cf. Figure 8) and that of referring expression (cf. Figure 9) especially conflict with the predictions of models in which predictability favours reduction: the subject antecedent is chosen less in the focussed subject than in the focussed object condition, but when it is chosen, more null pronouns are used to refer to it in the focussed object condition. Conversely, the object referent is chosen more in the focussed subject condition, and when it is chosen, it is referred to more often with heavier referring expressions such as overt pronouns and names in the focussed subject condition.

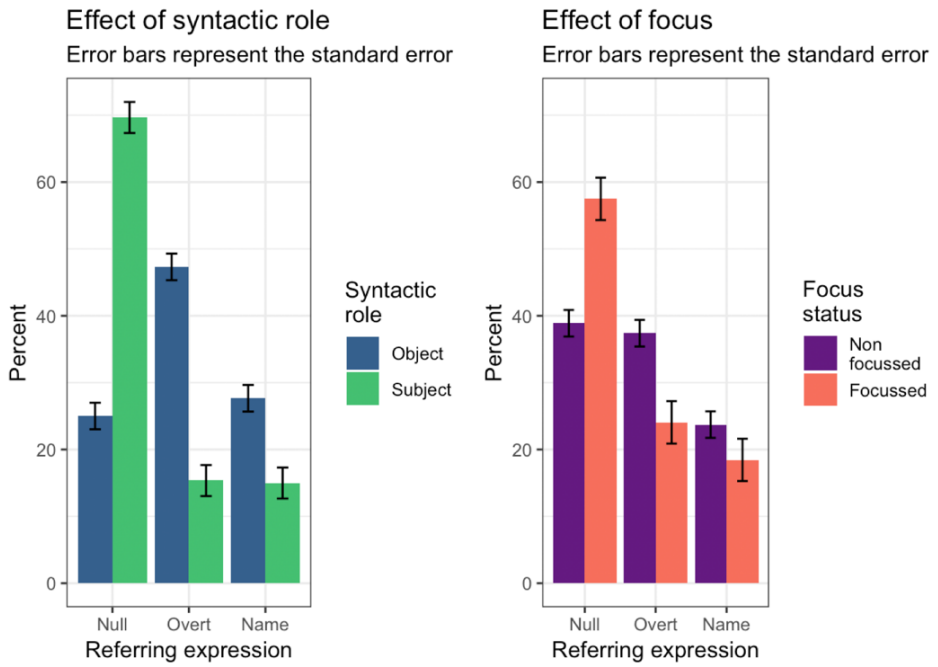

Figure 10. Effects of syntactic role and focus on referring expressions, collapsed across information structural condition.

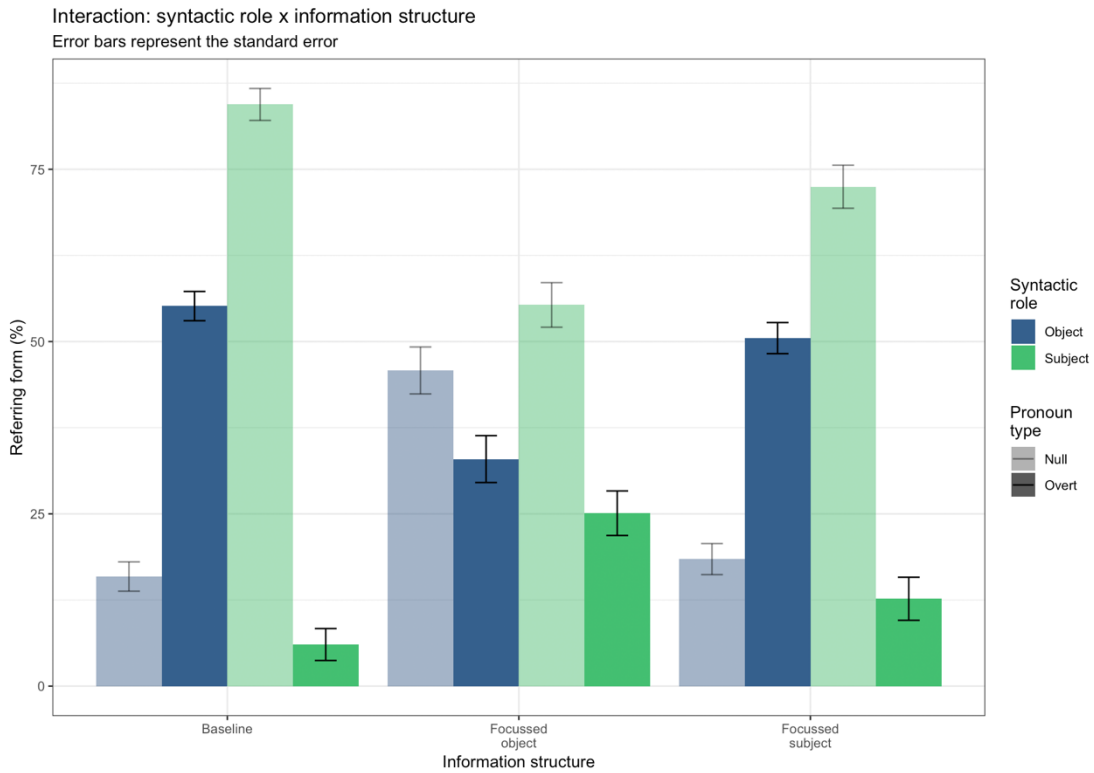

Figure 11. Effect of the interaction of syntactic role and information structure condition on referring expressions.

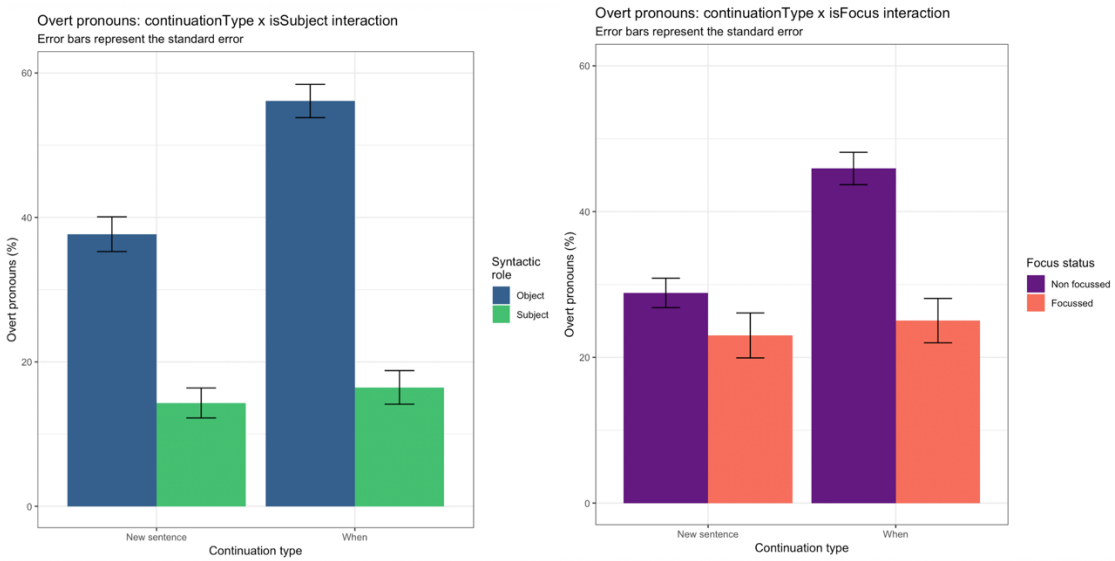

Figure 12. Effect of the interactions of continuation type with syntactic role and focus on referring expressions.

4.3.5.3 Analysis of coherence relations

The analysis of coherence relations excluded, in addition to ambiguous coreference, all cases of ambiguous coherence relations (i.e. cases in which the two annotators did not agree and cases which could be ascribed to more than one category). The data was further subdivided into the five most frequent coherence relations, namely Background, Explanation, Occasion, Result, and Violated Expectation. The two resemblance relations of Elaboration and Parallel were excluded because of their extreme scarcity in the data (9 occurrences each). The final data set includes 1311 continuations.

Making specific predictions about the choice of coherence relations is harder, but it is likely for sentence boundary to have a clear effect due to the experimental design using *quando* ("when") as a connective. The meaning of "when", in fact, is intuitively more compatible with occasions or explanations than with results or violated expectations. The effect of information structure is harder to pre-emptively gauge, although the marking of focus may act as a cue to signal an upcoming contrast, expressed for example as a violated expectation.

The effect of Information Structure and Continuation Type on coherence relations was also modelled with generalized linear mixed-effects models. Five models were fitted for each coherence relation.

The models were again defined following a step-down procedure starting from a maximal model with the interaction of the predictors and maximal random effect structure. All models eventually included Information Structure and Continuation Type without their interaction, except for the model for Background which also excluded Continuation Type as a main effect. All models were fitted including random intercepts for participant and prompt. This was the maximal random effect structure for all models but that for Violated Expectation, which could converge with a random slope for Information Structure by prompt but it was simplified for uniformity with the other models, as the complexity did not significantly improve the model ($p = 0.92$).

Bonferroni's correction was applied to the p-values to allow for the fact that the variable was modelled five times. The models' outcome is reported in Table 7.

The effects of Information Structure on the three affected coherence relations were confirmed by subdividing the data into couples of two of the three possible information structural conditions and fitting new generalised mixed-effects models. The results show that the baseline is significantly different from the focObj condition in its effect on isBackground, and both focussing conditions differ from the baseline for isExplanation and isOccasion:

- **Background relation sensitivity to Information Structure**: The only reliable difference was the lower rate of Background in the Baseline condition over the Object Focus condition ($p = 0.008$) (as compared to Baseline vs SubjectFocus: $p = 0.1$; and SubjectFocus vs ObjectFocus: $p = 0.28$);

Table 7. Coherence of the continuations: Estimated models' fixed effects.

Effect	Estimate	Std.Error	z value	$Pr(>\|z\|)$	
			isBackground		
(Intercept)	−1.42	0.25	−5.62	< 0.0002	***
focObj - base	0.5	0.19	2.63	0.009	*
focSub - focObj	−0.18	0.18	−0.98	0.33	
			isExplanation		
(Intercept)	−2.06	0.36	−5.74	< 0.0002	***
focObj - base	0.94	0.22	4.36	< 0.0002	***
focSub - focObj	−0.23	0.2	−1.11	0.27	
contType	0.61	0.18	3.51	0.0005	**
			isOccasion		
(Intercept)	−0.64	0.21	−3.1	0.0019	**
focObj - base	−0.7	0.17	−4.06	< 0.0002	***
focSub - focObj	0.06	0.17	0.37	0.71	
contType	0.73	0.14	5.13	< 0.0002	***
			isResult		
(Intercept)	−5.22	0.67	−7.78	< 0.0002	***
focObj - base	−0.92	0.39	−2.36	0.02	
focSub - focObj	0.88	0.39	2.26	0.02	
contType	−3.4	0.44	−7.8	< 0.0002	***
			isViolatedExpectation		
(Intercept)	−7.77	2.58	−3.01	< 0.0002	***
focObj - base	−1.93	2.95	−0.65	0.03	
focSub - focObj	2.29	3.09	0.74	0.02	
contType	−4.09	1.05	−3.9	< 0.0002	***

Significance thresholds with Bonferroni correction: $^{***}p < 0.0002$, $^{**}p < 0.002$, $^{*}p < 0.01$

- **Explanation relation sensitivity to Information Structure**: regarding the Explanation relation, two differences were reliable: the lower rate of Explanation in the Baseline condition over both the Object Focus condition (p < 0.001) and the Subject Focus condition (p < 0.001); SubjectFocus vs ObjectFocus did not reach significance (p = 0.25);
- **Occasion relation sensitivity to Information Structure**: with regard to the Occasion relation, two differences were reliable: the higher rate of Explanation in the Baseline condition over both the Object Focus condition (p < 0.001) and the Subject Focus condition (p < 0.001); SubjectFocus vs ObjectFocus did not reach significance (p = 0.74).

In the full models, Information Structure has an effect showing significant differences between focussed object and baseline structure, in which focObj produces more Background ($p = 0.009$) and Explanation ($p < 0.0002$) continuations, and fewer Occasion ($p < 0.0002$) continuations than the baseline. These effects are visualised in Figure 13.

The models show an effect of Continuation Type, with when continuations yielding more Explanation ($p = 0.0005$) and Occasion ($p < 0.0002$), and fewer Result and Violated Expectation continuations (both $p < 0.0002$). These effects are visualised in Figure 14.

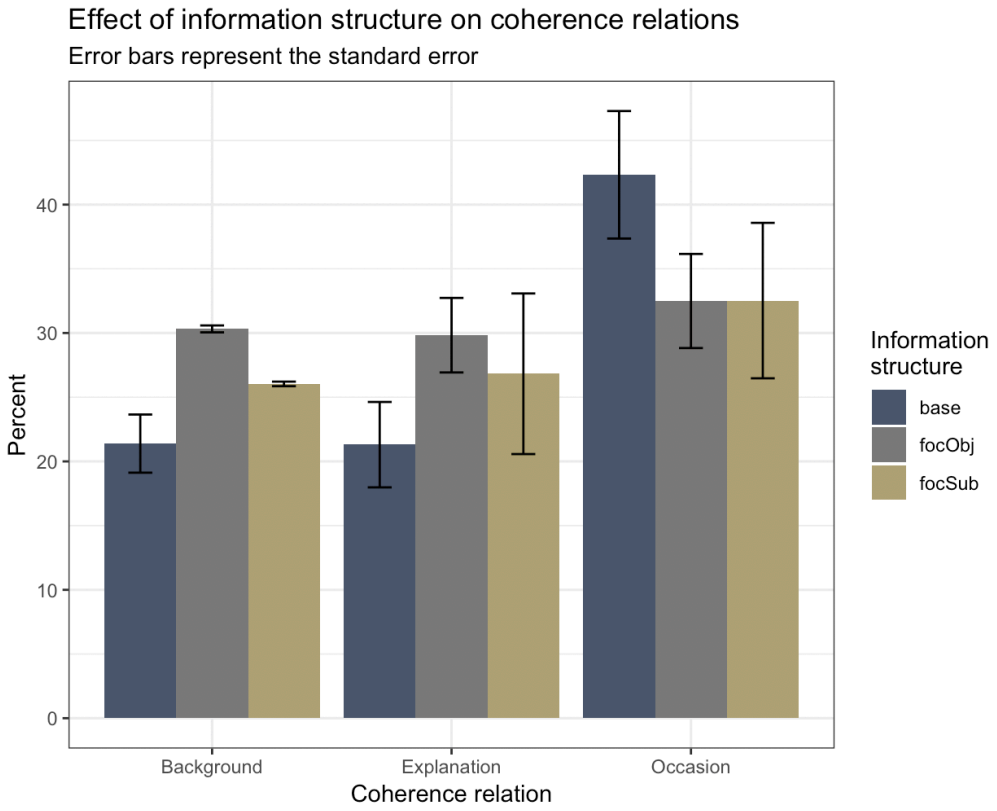

Figure 13. Effect of information structure on coherence relations.

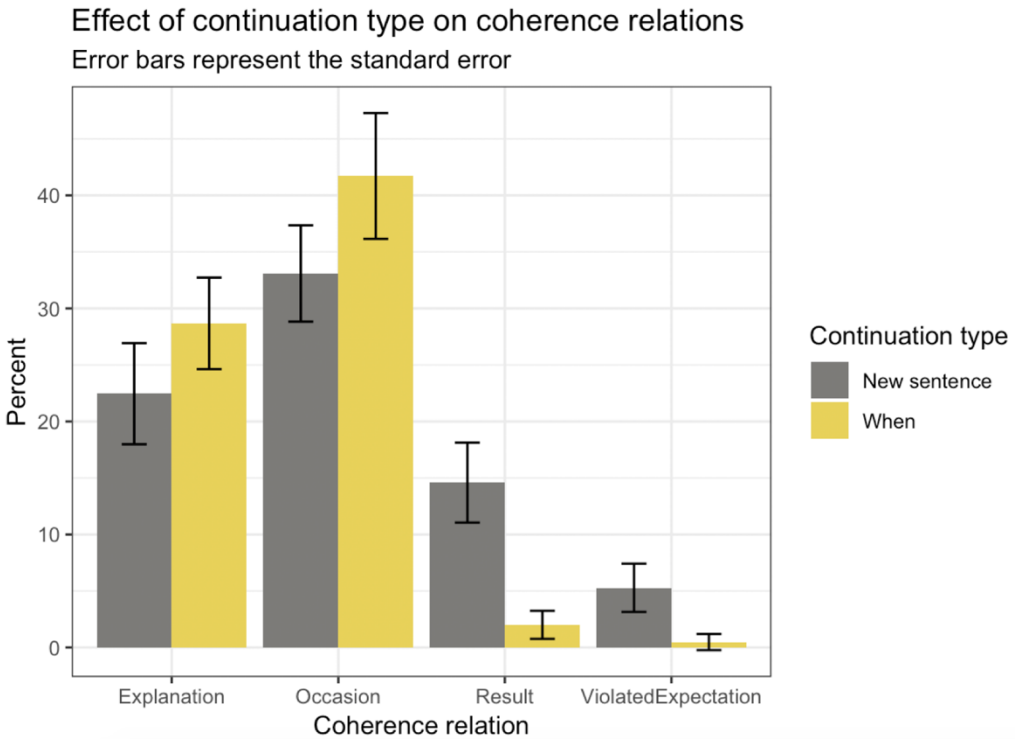

Figure 14. Effect of continuation type on coherence relations.

4.4 Discussion

4.4.1 Next mention

Following the predictions outlined in §4.3.1, sentence boundary on its own did not have a significant effect on antecedent selection. On the other hand, information structure showed an effect, most visible in the difference between the focussed subject and the focussed object conditions: starting from a general tendency to favour the object of the matrix clause for re-mention, referents are less chosen when they are focussed. This means that in clefts where the object is focussed, the subject antecedent will be favoured. The existence of an information structural effect (of any kind) on next-mention bias is not present in Kehler and Rohde (2018)'s Strong Hypothesis model, where information structure has an effect on referring expression choice only.

This dispreference for a focussed referent is in keeping with the results in de la Fuente and Hemforth (2013) on the interpretation of anaphoric null pronouns within-sentence in Spanish: in their study, when compared to a baseline, clefting the subject of a matrix

clause significantly increased the selection of the object for coreference and vice versa. This is also in accordance with "predictability ~ reduction" models and the claim that predictability is interrelated with topicality, rather than focality (e.g. Arnold, 2001; Tily and Piantadosi, 2009).

However, an interaction of information structure and continuation type was expected on the basis of a comparison between de la Fuente and Hemforth (2013)'s study and Cowles et al. (2007)'s study on English, in which, between sentences, focus was shown to increase the prominence of an antecedent and bias the antecedent selection accordingly. Although this prediction could be seen as weakly founded for a series of reasons – primarily cross-linguistic differences and the fact that relevant production, rather than interpretation, studies were not available – one could have assumed that an update of the common ground between clauses could have made a difference in the prominence of a focussed referent (see e.g. Zimmermann and Onea, 2011). Such an effect was not found, in that adding the interaction to the model did not improve its fit (nor did the interaction approach significance in a model including it).

This lack of interaction, and its significance in interpretation studies, confirms the lack of a production/interpretation mirror. Further research will be precious in assessing whether this pattern can be seen consistently across languages or a group of languages.

Finally, the choice of next mention was found to have an effect on the choice of referring expressions: this will be reviewed in more detail in §4.4.3.

4.4.2 Referring expressions used

Looking at the referring expressions produced shows at a first glance a preponderance of null subjects, making up 55% of the analysed forms. Overt pronouns in the "accusative" form and proper names account for respectively 27% and 18% of chosen referring expressions. Other referring expressions were produced, but they never reached 1% of all produced referring expressions, and were thus excluded.

These rates show first of all how the traditionally prescribed "nominative" form of the overt subject pronoun is no longer produced in Italian. Moreover, they show that proper

names make up a percentage of referring expressions that is too large to be left out of experimental settings, despite an often-mentioned "repeated name penalty" (e.g. de Carvalho Maia et al., 2017). Proper names are not considered in most interpretation studies on Italian and other null-subject languages, which target the null/overt binary. The present result problematises a notion of clean division of labour between the two pronominal forms as that posited by Carminati (2002), so that further research in interpretation studies will be needed to account for the function of other referring expressions.

4.4.3 Effects on referring expression selection

The production bias, unlike the next-mention bias, was influenced by both the information structure and clause boundary predictors, as well as by the choice of the antecedent a participant had made. It can be accepted that a writer (and a speaker), when choosing how to continue a sentence, will first choose who to talk about and what to say (i.e. next mention and coherence relations, the order of which cannot be assumed), and then choose how to express the continuation: that is, which referring expression to use.

Sentence boundary showed a clear effect whereby within-sentence continuations boosted the use of the minimal form, i.e. null pronouns, while between-sentence continuations increased the production of the strongest of the forms considered, i.e. proper names. No effect was seen on overt pronouns, making them a "buffer zone" between the most economical and the most informative expressions.

As in Rohde and Kehler (2014), information structure also had significant effects on referring expressions' selection. As with the next-mention bias, the biggest difference is to be seen between the baseline structure and the object-cleft sentence: focussing the object of the matrix clause makes participants use more null and fewer overt pronouns to refer to the object.

The choice of the selected antecedent also showed an effect on referring expression choice. This was modelled by looking at the effect of the syntactic [+subject] or the information structural [+focus] roles of the antecedent. The positive presence of both traits increased the use of null pronouns to refer to the antecedent and decreased that of both overt pronouns and names. Furthermore, this effect seems to be boosted in an interaction

with sentence boundary, where in "when" subclauses the proportion of overt pronouns used to refer to subjects and focussed referents is even smaller.

The information structural effect shows here a first contradiction that is difficult to account for with mirror accounts of prominence or salience: a focussed referent is less often chosen for re-mention, but when it is chosen, it is mentioned preferably through a minimal form. This hints at the fact that prominence is not as unitary as previously thought: a referent could be less salient than another for re-mention but still be referred to with a minimal form.

This is furthermore puzzling as the interaction of subjecthood and information structure overturns the effect of subjecthood: when compared to the baseline (information structurally unmarked) condition, in the focussed object condition the re-mention of a subject antecedent is carried out with fewer null and more overt pronouns and, conversely, an object antecedent is re-mentioned with more null and fewer overt pronouns.

4.4.4 Effects on coherence relations

Coherence establishment can be assumed to occur before the choice of a referring expression, around the same time as time as the choice of a next mention. Like for next mention, hence, the effect of the two predictors (continuation type and information structure) was modelled.

The effects of continuation type could be obvious because of the semantics of the word *quando* ("when") itself: clearly this connective is used more to talk about a time sequence, be it looking forward, like in Occasion (9a) or looking back, like in Explanation (9b); but the connective is less compatible with result relations like Result or Violated Expectation (9c).

9) **Examples of coherence relations extracted from the data**:

Michela helped Lorenzo solve the exercise when…

a. Occasion: *…she realised she didn't know how to do it.*

b. Explanation: *…he wanted to give up.*

c. ?Result: *…he got really angry.*

The influence of information structure is more difficult to account for. The effect, as elsewhere in the analysis, is most visible in the difference between the baseline and the focussed object conditions, and focussing the object decreases the occurrence of Occasion relations while increasing that of Background and Explanation.

The Occasion and Explanation relations are the most balanced in terms of subject/object split (respectively, 53% and 49% of subject choices). On the other hand, the Background relation is more object-biased (67% of object choices). This does not seem to lead to an explanation of the effect in which focussing the object referent, which makes it less chosen for next mention, affects the coherence relations in the same direction: in fact, in this case there should be fewer Background choices. Moreover, heavily object-biased relations like Violated Expectation and Result (respectively 88% and 94% object choices) may be predicted to be used less frequently, but this does not approach significance.

The effect of coherence relations on referring expressions was not modelled because of the high collinearity in the data. For example, Result relations are only expressed with a null pronoun in 4% of cases. However, as noted above, Result only appears with a subject antecedent in 6% of cases, and the use of null pronouns is more highly correlated with subject choices. Hence, the very low occurrence of null pronouns with Result relations could be due not to the relation per se, but to the choice of the antecedent.

The effect of information structure on the choice of a coherence relation is still reported as anecdotal evidence but its reasons, along with the effect that coherence establishment has on referring expressions (if any), need to be further studied in more targeted experiments, with fewer confounding variables.

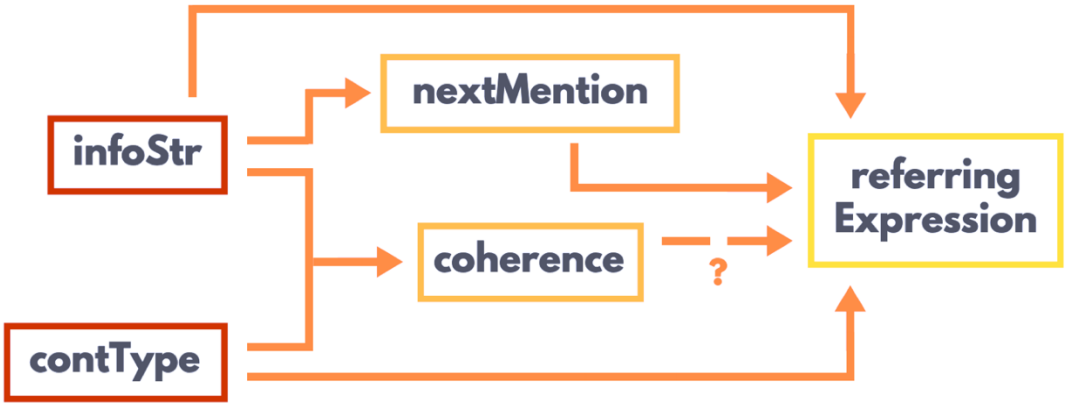

Figure 15. Representation of the effects from the manipulations in the prompts (red) to a phase in which participants chose "what to say" (orange), to the final choice they make on "how to say it" (yellow).

4.5 Conclusions

The reported study tests how manipulating information structure and continuation type influences the choice of a referent for next mention in a continuation experiment in Italian. Furthermore, it tests how this next mention, along with the previous variables, influences the choice of a referring expression. Finally, it starts to tackle the interaction of these elements with coherence relations.

The results of the study point to a very complex interaction between effects (Figure 15), whereby neither an interpretation-production mirror assumption nor a division of labour between a minimal and a marked form (null vs overt pronoun) hold. None of the models considered here and in Chapter 2 can completely account for the results of the study. Additionally, proper names were also chosen as referring expression along with pronouns. These results point to the fact that the dynamics between null and overt pronouns seen in Carminati (2002) and interpretation studies on Italian cannot simply be translated to the production of referring expressions.

Coherence relations were found to be affected not only by the connective used between the prompt and the continuation, but also by the information structural condition of the prompt. More research is needed to grasp their role in the production of anaphoric expressions and in patterns of referential choices.

This chapter contributes to assessing whether focussed referents are more prominent and, if so, what the implications of this prominence are. However, this raises the broader question of what parts of a sentence are more prominent. The next section (Chapter 5) will be concerned with this, studying whether events as a whole or the entities implicated in them are more prominent, and whether this relates to referring expressions. Two story continuation studies will be presented (respectively on English and a comparative one on English, French, German, Italian and Spanish) wherein referring expressions are manipulated alongside the complexity of the antecedent, using verb status in the causative-inchoative alternation as a proxy to make a second implicit entity available or not. The results will show patterns that are consistent cross-linguistically and can be related to some of the theories overviewed in Chapter 2, and in particular to Ariel's (1988, 1990) Accessibility Theory (§2.3).

Chapter 5 Event and entity coreference

What parts of a sentence are more prominent?

5.1 Overview

In the second part of this thesis, the focus zooms in from prominence and its repercussions at sentence level to the prominence of different components of meaning, first, here, at the level of event structure within a sentence, then (Chapter 6) at the level of a single complex entity. At the event structure level, two different components that can be observed are events and the entities that are involved in them.

Event coreference is much rarer than entity coreference (cf. §5.3.1 and §5.3.5). Moreover, events are different from entities in that they are more complex and usually not directly introduced (Hedberg et al., 2007). Both personal and demonstrative pronouns, among other referring expressions, can be used to refer to either entities (1a) or events (1b), without a clear division of labour:

1) The bomb planted by the terrorists failed to explode.
    a. It was faulty.
    b. It was lucky.

This chapter examines the question of whether these factors make events more or less retrievable than entities when it comes to selection of antecedent and of referring expression. To answer these questions, two story continuation experiments are presented: the first on English (§5.2); the second on five languages: English, French, German, Italian and Spanish (§5.3), to compare the results crosslinguistically and determine whether the effects of prominence and distributions of referring expressions pattern similarly across languages (pointing to an across-the-board cognitive strategy) or rather if they are language-specific and dependent on the typology or grammatical attributes of the language.

Personal and demonstrative pronouns are used as prompts to elicit continuations that may refer either to the entity presented in the previous sentence, or to the whole event in

which the entity is involved. Events are both less commonly referred to and more complex; that being the case, Accessibility (Ariel, 1990; see §2.3) would predict that heavier referring expressions would preferentially refer to them, and lighter referring expressions would bias coreference towards the simpler entities: in this case, a bias is predicted for demonstratives to corefer with events, and for personal pronouns to corefer with entities, as research has shown to be the case for both processing and production of anaphoric forms (Çokal et al., 2018). This bias is shown examples (2a) and (2b), extracted from the data of the experiment in §5.3, respectively show:

2) The snow that was covering the fields was melting down.
    a. This allowed us to see the grass and flowers emerging.
    b. It had turned into slush and mud.

The studies manipulate features that are known to have effects on reference and event status. Other than the referring expressions, verb aspect and category in the causative-inchoative alternation are manipulated: these are used as proxies to represent an event as either ongoing or completed, and to have two levels of complexity for the event varying the availability of an additional implicit argument of the verb. The aspect of verbs has been shown to influence the prominence of an event and, with it, coreference patterns: end-state (perfective) events have been shown to be more easily retrievable (Moens and Steedmann, 1988); however, ongoing events have also been shown to be highly activated because of their unfinished state (Ferretti et al. 2009).

The alternation of verbs, on the other hand, does not concern the state of an event but rather its structure: alternating verbs make an additional referent available either explicitly or implicitly, as shown in (3), where in (3a) the entity "sun" is explicit as the agent in the event, while in (3b) the additional entity is implicit:

3) Causative-inchoative alternation:
    a. Causative construction: *The sun melted the snow*.
    b. Inchoative construction: *The snow melted*.

In verbs that do not allow for the alternation seen in (3), e.g. "die" (cf. *My phone died* / **Water damage died my phone*), the single available entity may be more prominent because it has no competition from other entities, as can be inferred from Centering Theory (Grosz et al., 1995; §2.3) and from the Structure Building Framework (Gernsbacher, 1990; §2.4). Such shifts in prominence caused by competing entities may affect the referring expressions chosen for anaphoric reference, as has been shown even in the absence of ambiguity (Arnold and Griffin, 2007). Alternation is a gradable phenomenon, with some verbs having rare causative uses and standing in between categories, and this division may be different across languages.

The overall findings show crosslinguistically consistent results: heavier referring expressions, such as demonstratives, are used for events rather than entities. Moreover, increasing the complexity of an event by adding an implicit entity via an alternating verb also elicits more event coreference. This can be explained by the increased competition to the explicit entity that the additional entity causes, as predicted by Centering Theory (Grosz et al., 1995). Finally, presenting an event as completed also makes participants choose it for continuation more often (although not to the extent that event complexity does).

Both studies presented in this chapter are result of a collaboration within a research group composed of Hannah Rohde, Sharid Loáiciga, Christian Hardmeier, and myself. This is the same group that also produced the studies in Chapter 6, and represents a collaboration that emphasises the crosslinguistic study of coreference across the domains of Linguistics and Computer Science. The first article (§5.2) was included in the *Workshop on Computational Models of Reference, Anaphora and Coreference (CRAC)*; the second article (§5.3) has been submitted to *Dialogue & Discourse* and is currently being reviewed. In both studies I took part in the design, testing and analysis phases, and the writing too was produced as a group effort.

The first article is presented verbatim (with the exception of numbering) in its published form (§5.2), the second makes small changes to its submitted form (§5.3).

5.2 Event versus entity co-reference: Effects of context and form of referring expression[14]

5.2.1 Introduction

A challenge in discourse interpretation is the resolution of referring expressions, particularly those whose meaning is compatible with many potential antecedents. To take an example like (4), a passage may introduce a number of entities and situations that a subsequent sentence might refer to.

4) Everybody who is involved with this debate has been struggling over me and my personality. [ParCorFull]

For a sentence following (4), certain expressions would be resolved unambiguously to a unique entity (e.g., to the speaker for a 1st person singular pronoun *I*) or would easily be linked to the only compatible referent in the context (e.g., to the group of relevant individuals described as *Everybody* for a 3rd person plural pronoun *They*). Other expressions are compatible with more than one entity (e.g., the debate or the personality for a pronoun *It*) and therefore create potential ambiguity. Making matters worse, the antecedent of some expressions can be either an entity or something more abstract: an event or situation or idea. Such expressions include personal pronouns like *It* and demonstrative pronouns like *This*/*That*.

Given the complexity of identifying a set of candidate abstract antecedents in a given context and then determining whether a particular expression is re-mentioning one of those abstract antecedents or a more concrete entity, many co-reference systems focus only on nominal antecedents (e.g., BART, Stanford's sieve-based, HOTCoref (Versley and Björkelund, 2015)). However, event instances are also referential.[15]

[14] Loáiciga, S., Bevacqua, L., Rohde, H., & Hardmeier, C. (2018). Event versus entity co-reference: Effects of context and form of referring expression. In *Proceedings of the Workshop on Computational Models of Reference, Anaphora, and Coreference (CRAC)*. Available at https://www.aclweb.org/anthology/W18-0711.

[15] Here we call *event* what is more commonly known as *abstract anaphora* (cf. Dipper and Zinsmeister (2010); Nedoluzhko and Lapshinova-Koltunski (2016)). We take as an *event* any non-nominal relationship for the pronouns *It* and *This* and a *textual* antecedent in the form of a text span of variable length (e.g., a word, a clause, several sentences). *Textual* means that anaphoric relations for which some type of inference is necessary are not included, e.g., bridging or extra-textual reference. The term *event reference* is founded upon

This paper asks when and to what degree event instances serve as antecedents when a competing entity referent is also available. The goal is to model human choices as a baseline to inform co-reference systems. We report two psycholinguistic studies that use a story-continuation task to measure participants' resolution of pronouns *It* and *This*.

Improving our understanding of the interpretation of the "difficult" anaphoric cases is a step towards better anaphora and co-reference systems. It has been noted that current systems struggle to identify this type of reference and that anaphoricity determiners have poor performance (Heinzerling et al., 2017). *It*, *This* and *That* are also frequent in dialogue data for which co-reference systems' performance is particularly low (Eckert and Strube, 2000; Müller, 2007). In addition, pronoun function is relevant to the evaluation of machine translation systems since different functions entail different translations according to the constraints of the language pair and can thus affect performance (Guillou, 2016).

5.2.2 Related work

Both corpus-based and psycholinguistics works on the interpretation of anaphoric expressions concentrate on the identification of the antecedents of nominal expressions. Abstract anaphora – anaphora that involves reference to abstract entities such as events or states (Asher, 1993) – is much less studied from both fields, as evidenced by the little amount of annotated data available (Dipper and Zinsmeister, 2010; Poesio, 2016).

Corpus-based studies of pronouns are often done in relationship to the texts on which co-reference resolution systems will be trained and tested. With the clear aim to improve precision, the authors of these systems have an interest in quantifying "non-anaphoric" pronouns for preventing their resolution. We know for instance, that about 5% of referential pronouns and 71% of demonstratives in dialogue data refer to events (Müller, 2007; Poesio, 2016), whereas about 3% of referential *it* pronouns in written text of various genres refer to events (Evans, 2001).

Webber (1986), and we set on the name *event* for the sake of consistency with the annotation in the corpus used in the second study presented here.

In psycholinguistic research, on the other hand, the focus has been on using theoretical constructs of complexity, salience, and focus to capture co-reference patterns. The demonstratives *This* and *That* have been grouped together, assuming that they behave in the same manner, but potentially differently from *It*. Brown-Schmidt et al. (2005) analyze *It* vs *That* and report a preference for *That* if what is referred to is a composite (e.g., *I'll have the hamburger and fries. I'll have that, too.*), independent of other metrics of the salience of the referent. Building on the Centering co-reference model (Grosz et al., 1995), Passonneau (1989) analyzes intra-sentential instances of *It* vs *That* with an explicit NP antecedent. She reports that *It* is used to refer to the center (most often the subject), whereas *That* favors non-centers.

Corpus-based studies offer insights about language use, since the written texts they are based on are, after all, natural passages. They offer better estimates for building systems that will be used on those texts. Corpus-based studies, on the other hand, do not offer any explanation as to why a particular item follows a certain distribution, and they grant little control over the confounding variables responsible for that distribution. In this respect, psycholinguistics studies provide more suitable methods for capturing the cognitive processes behind naturally occurring phenomena. We therefore start the next section with a study using constructed passages to allow for careful control over format and content.

5.2.3 Study 1: Constructed passages

A story-continuation experiment was conducted to establish a baseline rate at which participants assign *It*/*This* pronouns to entity vs event antecedents. By varying a property of the context sentence, we test how malleable the two pronouns' respective co-reference preferences are. A 2x2 design manipulated the context sentence (alternating/non-alternating verb) and the pronoun prompt (*It*/*This*, as in (5)-(6)).

5) The train from the Highlands arrived promptly. It/This ____
6) The balloon with the red hearts popped noiselessly. It/This ____

The availability of entities for anaphoric resolution is dependent on the argument structure of the previous predicates. Alternating verbs can have an intransitive as well as a tran-

sitive use: the first usually describes a change of state (7-a), and the latter specifies, in subject position, which entity brought on the change (7-b). Conversely, non-alternating verbs do not allow a transitive use (8).[16]

7) a. The snow melted.
   b. The heat melted the snow.
8) a. The battery died.
   b. *The heat died the battery.

Manipulating the verb in the context sentence affects the argument realization options associated with the predicate: Non-alternating verbs like *arrive* permit only a single realization with the entity that arrives always in subject position; alternating verbs like *pop* are compatible with realizations where the entity that pops appears in subject position or object position. For alternating verbs, an explicit agent entity can be introduced (*I popped the balloon*) or left implicit, as in (6).

One hypothesis is that alternating verbs could make available an additional (implicit) agent who might provide more entity co-reference opportunities and thereby increase entity co-reference and reduce event co-reference. Another hypothesis is that non-alternating verbs may make salient one single (explicit) entity by eliminating competition from other (implicit) entities and thereby yield more entity co-reference and less event co-reference. The existence of an external, unspecified argument in the syntax of alternating verbs is still controversial (Embick, 2004; Schäfer, 2009), but the cognitive accessibility of a possible agentive entity arises from the very fact that the causative alternation exists.

Although differences have been observed between the use of proximal and distal demonstratives *this* and *that* (Çokal et al., 2014), we targeted only one demonstrative pronoun in order to simplify the design. This is in keeping with observations about the functional

[16] Jespersen (1927) collects verbs undergoing alternation in a "move and change class". They have also been referred to as respectively *causative* and *anticausative* (or *inchoative*) verbs (Schäfer, 2009); the phenomenon has also been studied as "causative-inchoative alternation" (Haspelmath, 1993).

grouping of a number of pronouns (zeros, demonstratives, and personal pronouns) when used deictically (Webber, 1990).

5.2.3.1 Materials

The 24 experimental items consisted of a context sentence and a pronoun prompt, as in (5)-(6). Participants saw all items, with either *It* or *This*. Subject NPs were modified (8 nouns with pre-nominal adjectives, 8 nouns with post-nominal prepositional phrases, 8 nouns with post-nominal relative clauses). The verb used an adverbial or particle predicate (roughly half alternating, half non-alternating). The head of the subject NP was always the only singular entity, with any other mentioned entities being incompatible with 3rd person singular co-reference (e.g., *we* or *the red hearts*).[17]

The 24 experimental items were interleaved with 40 filler items. These included 20 passages with a context sentence mentioning one or two entities, followed by a discourse adverbial prompt (e.g., *As a result*, *Then*), 16 passages for an unrelated experiment involving mentions of companies and other organisations, and 4 catch trials with an obvious correct response (e.g., *Caleb did all the cooking for the BBQ even though he hates BBQ. He prefers mac 'n* ____).

5.2.3.2 Participants

Twenty-seven monolingual English-speaking participants aged 19-63 (mean age 36, σ = 11.2; 15 male) were recruited from Amazon's Mechanical Turk (Munro et al., 2010; Gibson et al., 2011) and received $4 for an estimated 30-minute task.

5.2.3.3 Procedure

Continuations were collected via a web-based interface that participants could access from their own computer. Each item was presented on a page by itself with a text box for participants to use for writing their continuation.

[17] The data to reproduce our experiments and the full models can be found on https://github.com/sharid-loaiciga/event_vs_entity.

5.2.3.4 Annotation and analysis

Continuations for experimental items were annotated for type of co-reference (entity vs event). The four authors of this paper shared the annotation such that all target continuations were coded by two annotators. To be conservative, annotators were blind to the *It*/*This* prompt condition and agreed to err on the side of annotating a pronoun as ambiguous if the pronoun could be interpreted plausibly as coreferential with an event or an entity (e.g., *The brand new siren sounded loud. [omitted pronoun] startled some people*).

Using mixed-effects logistic regression, we modeled the binary outcome of entity or event co-reference with fixed effects for prompt type, verb class, and their interaction, with maximal random effects structure when supported by the data (Barr et al., 2013). Where a model did not converge, we removed random correlations. All factors were centered. Reported p-values are from glmer model output using the lme4 package (Bates et al., 2015a) in R (R Development Core Team, 2008).

5.2.3.5 Results

Of the 626 total continuations, we excluded 128 that were judged by one or more annotators to be ambiguous (or for which the annotators gave conflicting annotations) as well as 55 that used the prompt in another way (e.g., *This noun*). This left 443 continuations with either entity or event co-reference. Note that at the analysis stage, 2 of the 24 verbs were re-classified as alternating verbs, shifting the original even split between alternating/non-alternating verbs. However, glmer models are understood to be robust against datasets that are not perfectly balanced.

The results (see Figure 16) show a strong, but not categorical, bias to use *It* to refer to entities and *This* to refer to events. In addition, verb type impacts co-reference, whereby verbs that permit alternations yield more event co-reference than non-alternating verbs. This is in keeping with our second hypothesis that the salience of the single argument of non-alternating verbs may have attracted more entity co-reference.

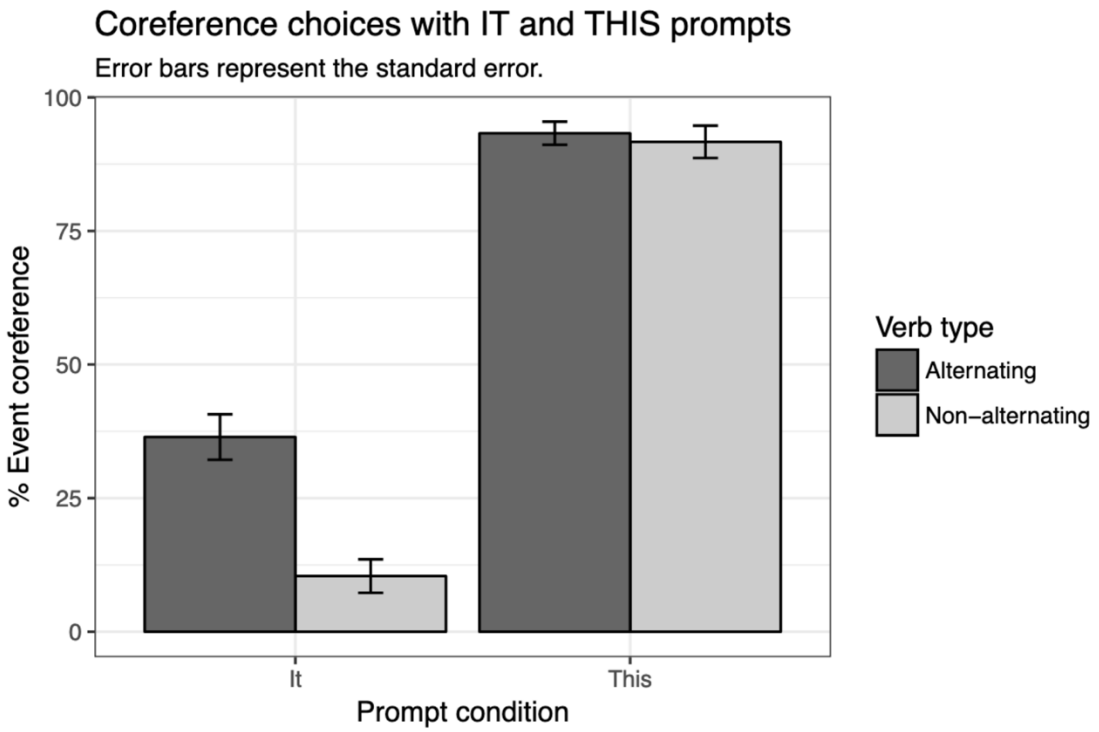

Figure 16. Study 1 results by prompt and verb type.

The prompt type × verb type model of co-reference choice confirms a main effect of prompt type ($\beta = 5.100$, $p < 0.001$) and a main effect of verb type ($\beta = 1.437$, $p < 0.05$). There was no prompt × verb type interaction ($\beta = -1.350$, $p = 0.22$).[18]

5.2.4 Study 2: Corpus passages

5.2.4.1 Materials

The 48 target passages are minimally edited sentences extracted from the ParCorFull corpus (Lapshinova-Koltunski et al., 2018). This is a German-English parallel corpus annotated with full co-reference. Although the corpus is designed for nominal co-reference, it includes annotations of two types of antecedents: entities and events. Entities can be either pronouns or NPs, whereas events can be VPs, clauses or a set of clauses.

[18] Inspection of Figure 16 suggests a possible interaction whereby the effect of verb type looks stronger in the *It* condition than in the *This* condition. The lack of a significant interaction in the model may reflect the fact that the co-reference rate for non-alternating verbs in the *This* condition is already near ceiling and there may be little room for (measuring) a further increase.

ParCorFull includes texts from TED talks transcripts and also newswire data.[19] Since pronouns are generally more frequent in the TED talks genre than news, we concentrated on this portion of the corpus only. Twelve examples of each *It-entity*, *It-event*, *This-entity*, and *This-event* were selected. In comparison to the sentences from Study 1, the corpus sentences were relatively long; therefore, simplified or shortened versions were used.

Additionally, the target passages were interleaved with 52 filler items. From these, 24 were extracted from ParCorFull sentences with no annotation and a continuation starting with an adverbial expression was prompted (e.g., *The encyclopedia business in the days of leatherbound books was basically a distribution business. Eventually,* ____). 24 other fillers were extracted from the OntoNotes corpus (Pradhan et al., 2013) for a dataset for an unrelated experiment involving mentions of companies and other organisations, as in Study 1. A final 4 fillers repeated the catch trials from Study 1.

5.2.4.2 Participants

Nineteen monolingual English-speaking participants aged 23-44 (mean age 30, σ = 6.5; 13 male) were recruited from Amazon's Mechanical Turk and received $7 for an estimated 50-minute task.

5.2.4.3 Procedure, annotation and analysis

The procedure was identical to that in Study 1. The annotation followed that described for Study 1. As an illustration, example (9) shows a passage whose original co-reference relation was one between an *it* pronoun and an entity antecedent. The continuations in (10) were annotated as event co-reference (10a), entity co-reference (10b), and no co-reference when the *It* prompt was classed as being used pleonastically (10c).

9) You carry a phone. It knows where you are. [original co-reference: entity~*it*]

10) a. You carry a phone. This is something that just about everyone does these days.

b. You carry a phone. It is capable of connecting you to others and

[19] Specifically, the ParCorFull corpus includes the datasets used in the ParCor corpus (Guillou et al., 2014), the DiscoMT workshop (Hardmeier et al., 2016) and the test sets from the WMT 2017 shared task (Bojar et al., 2017).

the world around you.

c. You carry a phone. It wouldn't hurt you to call once in a while.

The binary outcome of entity/event co-reference was again modeled with a logistic regression. We included fixed effects for prompt type, original passage co-reference (entity/event), original passage referring expression (*it*/*this*), and the 2-way and 3-way interactions. All factors were centered.

5.2.4.4 Results

Of the 788 total continuations, we excluded 94 that were judged by one or more annotators to be ambiguous (or for which the annotators gave conflicting annotations) as well as 98 that used the prompt in another way (e.g., *This noun*). This left 596 continuations with either entity or event co-reference.

The results (see Figure 17) follow those of Study 1 for the prompt manipulation: Event co-reference is higher with *This* than *It*. Event co-reference further increases when the original passage contained event co-reference. The model (prompt type × original passage type × original passage pronoun) confirms a main effect of prompt type ($\beta = 2.529$, $p < 0.001$) and a main effect of original passage type ($\beta = 3.053$, $p < 0.001$), with no effect of original pronoun or any significant interactions.

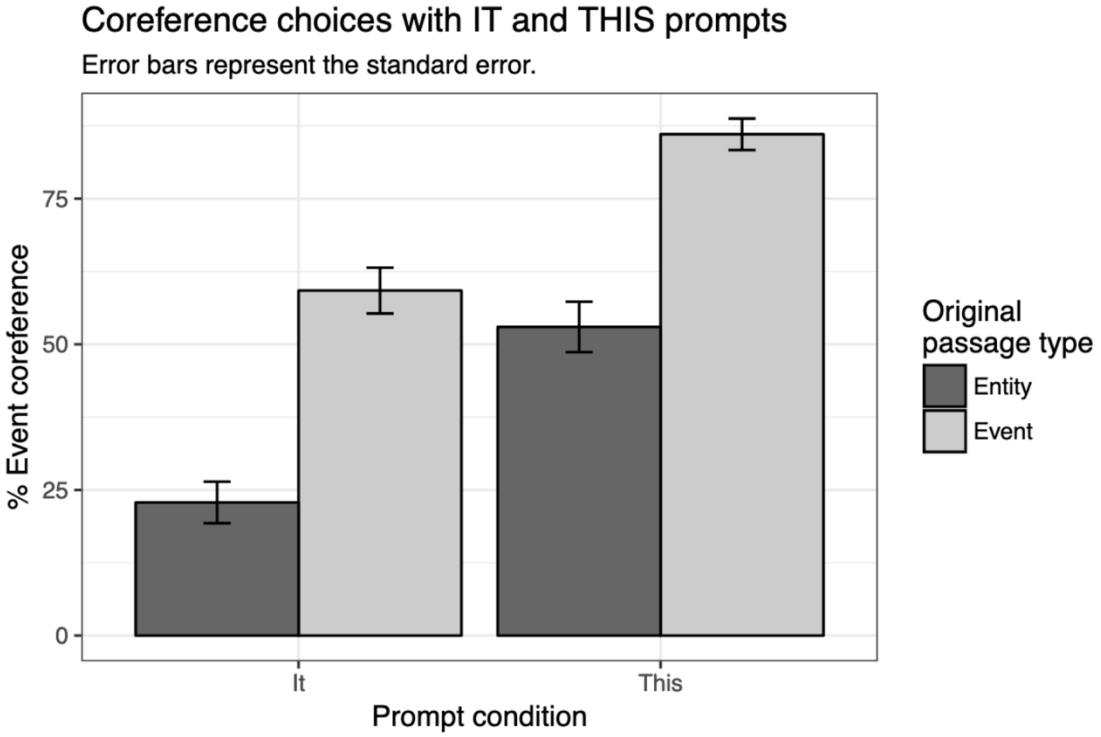

Figure 17. Study 2 results by prompt and original co-reference (collapsing over original *it*/*this* pronoun type)

5.2.5 Discussion

The two studies show divergent co-reference distributions for the personal pronoun *It* and the demonstrative *This*: a bias towards entity co-reference for *It* and a bias for event co-reference for *This*. As far as we know, this pattern has been proposed (Dipper and Zinsmeister, 2010), but not properly measured. Given the oft-assumed division of labor between these two pronouns, what is notable is their flexibility. Neither form was found to be used categorically in Study 1 or Study 2.

Interestingly, the study with the constructed passages showed that verbs which permit an agent alternation as either an implicit or explicit argument are more prone to trigger an event co-referent than an entity one. This finding is potentially useful as an additional feature for anaphoricity detection or event mention identification in co-reference resolution systems.

Furthermore, we saw a bias towards event co-reference for the corpus passages in Study 2 that were known to have yielded event co-reference in their original passages. This suggests that there are properties of the context sentence that may make salient an event over an entity. If there are event-favoring properties of the context sentence that human participants are sensitive to, it is a tractable task to build automatic classifiers that learn to recognize such properties. This supports the idea that the task of differentiating anaphoric and pleonastic instances of *It* (Evans, 2001; Boyd et al., 2005; Bergsma and Yarowsky, 2011; Lee et al., 2016; Loáiciga et al., 2017) could potentially improve performance.

Although presumably (machine) learnable, the question of what exactly constitutes an event remains unanswered. A number of ambiguous examples which were excluded from our analysis included entities that are close to their entailed event (e.g., *The bomb that the arsonists had planted exploded violently*) or that were very abstract (e.g., *The greatest opportunity materialized unexpectedly. It/This was almost like magic.*).

5.2.6 Conclusions and future work

This paper reports an investigation on abstract anaphora. Specifically, two studies targeted the ambiguity that occurs when entity and event antecedents are available for the pronouns *It* and *This*. A clear pattern emerged whereby *It* favors entity co-reference and *This* favors event co-reference. This pattern is also affected by the number of arguments that the main verb can take. Although further investigation is needed regarding the properties of events, their salience, and the gray area between events and entities, our results take a first step towards disentangling the behavior of less well-understood anaphoric relations.

5.3 Event and entity coreference across five languages: Effects of context and referring expression[20]

5.3.1 Introduction

Work in coreference has focussed primarily on entity coreference, typically between third person personal pronouns and human antecedents. The use of anaphors to refer to non-human entities has been less studied, especially when the reference is to antecedents that are not entities, such as events or full clauses. Non-entity coreference relations present both formal and practical challenges. If the antecedent can be any of the available previously mentioned entities, events, propositions or even entire rhetorical arguments, the number of candidate antecedents becomes potentially very large. In addition, the anaphoric forms that speakers use for different kinds of coreference are not mutually exclusive: sentences (11) and (12) show the demonstrative *this* referring back to respectively an entity (the wine) and an event (the successful aging of the wine).[21]

11) The Northern Italian wine aged well. This was one of the best wines I'd ever tasted.
12) The Northern Italian wine aged well. This made the wine more marketable.

This ambiguity has been the subject of recent work on the coreference system in English, with evidence of a split between comprehenders' preferential interpretation of personal versus demonstrative pronouns (Loáiciga et al., 2018). Specifically, personal pronouns are preferred when referring to entities, and demonstratives when referring to other more complex or abstract antecedents (see §5.3.2.1). Demonstratives have in fact been identified as the expressions speakers use to signal a prominence shift, particularly to change the discourse topic (Givón, 1983) and to reject the most prominent antecedent as a possible antecedent (Comrie, 1997).

Different languages use different referring expressions, ordering the elements in a given language's pronominal system along a prominence scale: for example, while in English

[20] Submitted to *Dialogue & Discourse* under the title *Event and Entity Coreference Across Five Languages: Effects of Context and Referring Expression* and with authors Bevacqua, L., Loáiciga, S., Rohde, H., and Hardmeier, C. First authorship shared by Bevacqua and Loáiciga. Currently in reviewing stage.
[21] All examples are taken from the experiments described in §5.3.3.

personal pronouns encode highly accessible antecedents, in languages with widespread use of zero anaphors such as Italian or Spanish, personal pronouns will preferentially be used to refer to less accessible referents. Moreover, less prominent antecedents will need more specific types of referring expressions in order to be retrieved, whereas more prominent antecedents can be referenced with a wider variety of referring expressions (von Heusinger and Schumacher, 2019).

In addition, when processing coreference, comprehenders are shown to be sensitive to a wide range of features related to structural and meaning-driven properties of the passage (see reviews such as Greene et al., 1992; Carlson, 2003; Rohde, 2019). Those findings include a large body of work in psycholinguistics, much of which has emphasized English (with notable exceptions of course, e.g., Kaiser and Trueswell, 2004; Mayol, 2018). Here we pick up on two English findings from a psycholinguistic study: antecedent complexity (number of arguments) and aspect, two verb-driven cues that signal how an event is being portrayed.

The accessibility of a target antecedent decreases with a higher number of possible competitors (see e.g. Centering Theory: Grosz et al., 1995). One way to operationalise antecedent count and complexity is to consider the category of a verb within the causative-inchoative alternation (Haspelmath, 1993) can then be used as a proxy for antecedent choice complexity. Alternating verbs admit both intransitive and transitive uses (e.g., *The snow melted* / *The sun melted the snow*), corresponding to inchoative and causative interpretations, respectively. This alternation is not present in all verbs (e.g., *The battery died* / **Frequent use died the battery*). Jespersen (1927) classes verbs that can undergo this alternation as "move and change verbs'", as they often describe changes in state or movement: the transitive use, then, makes available the implicit entity that brought on the change. Alternating verbs, by offering this additional implicit entity to the pool of coreference possibilities even when the verb appears in the intransitive, have been shown to increase event coreference in English (Loáiciga et al., 2018); for non-alternating verbs, the sole entity available in the sentence seemingly becomes relatively more salient by not having competition from other implicit entities. However, this is a gradable phenomenon, in that some verbs position in between categories: even verbs which are most firmly in the non-

alternating category can have rare or idiomatic transitive uses (e.g. internal objects, *to die a cruel death*). Languages may differ in how clean-cut the divide between alternating and non-alternating verbs is, and in how sensitive coreference is to this distinction.

Verb aspect used as a cue to the portrayal of an event, on the other hand, may influence a comprehender's perception of the concreteness of the event or its availability to be referred to: events which are portrayed as completed might have a salient end-state and, thus, be more encapsulated and retrievable (Moens and Steedman, 1988) or, on the contrary, the ongoing state of an event may activate it and, with it, its properties and world knowledge (Ferretti et al., 2001, 2007, 2009).

The relevance of the studies we present here is underscored not only by the lack of work on abstract anaphora and the possibility of making crosslingual extensions to the limited prior work on this topic, but also by the similar questions raised in coreference-annotated corpora that inform computational systems for coreference resolution. Existing resources in the field of coreference resolution also reflect the difficulty in the treatment of pronouns. For example, given the complexity involved in identifying non-nominal anaphora, and its relatively low frequency, most annotation efforts and hence most coreference systems focus on nominal anaphora. The OntoNotes corpus (Pradhan and Xue, 2009), for instance, although the largest in the community, does not include annotations of event entities coreferential with pronouns. Corpora including events exist, but they are smaller in size and less frequently used. The ParCorFull corpus (Lapshinova-Koltunski et al., 2018), for instance, includes the annotation of event anaphora in its guidelines, but event references only constitute 12.5% of the total markables (annotated elements) in English. ParCorFull is a parallel corpus that also includes annotations in German and French. In German, event annotations amount to 11%; in French the proportion is even lower, however it should be noted that the lower number of documents for this part of the corpus makes the estimate less reliable.

Improving our understanding of the interpretation of event anaphora and non-nominal anaphora in general is a step towards better anaphora and coreference systems (Poesio et al., 2015). Current systems struggle to identify and resolve this type of reference simply

because they have not been trained to do so (Heinzerling et al., 2017), as annotated corpora focus on nominal anaphora. Annotated corpora are crucial for system development and large scale linguistic analysis, but equally important is a sound theoretical understanding of the phenomenon one wishes to annotate.

This paper presents work on these different types of coreference via a comparison across five languages (English, French, German, Italian, and Spanish), which are broadly related typologically but which differ in their use of grammatical gender and case and their pronominal systems, most notably the availability of a null pronoun.

The paper asks when and to what degree event instances serve as antecedents when a competing entity referent is also available. The goal is to model human choices in five different language settings to investigate the similarities and differences that arise as a product of the structural differences in the languages, and to shed light on the relationship between event and nominal anaphora in order to inform future coreference annotation efforts and coreference systems.

We report five story continuation studies targeting participants' resolution of pronominal forms *It* vs *This* in English, *Il* vs *Cela* vs *C'est* in French, *Es* vs *Das* vs *Dies* in German, *Ciò* vs *Questo* vs zero pronoun in Italian, and *Esto* vs *Este* vs zero pronoun in Spanish.

Regarding antecedent complexity and aspectual status, we test the impact of (i) the availability of an additional implicit argument to the verb, namely the effect of verbs presenting a causative-inchoative alternation and (ii) the aspectual difference of portraying of an event as completed or ongoing. Regarding anaphoric forms, we test (iii) whether event coreference is inferred more for demonstratives over personal pronouns and (iv) whether this distinction is categorical, with a clear division of labour, or gradable.

Our results show consistent and gradable biases: participants encountering a personal pronoun are more likely to continue the sentence writing about the entity, while demonstratives prompt more event continuations across all five languages. The complexity of the antecedent, manipulated through the proxy of alternation, increases event coreference, though this is significant in only four of the five languages studied. The portrayal of

an event as completed, on the other hand, seems to have a much smaller role in influencing coreference. We speculate how the varying aspectual systems of the five languages may undermine this effect. Lastly, we report corpus frequencies comparing the use of the same referring expressions to refer to entity and event antecedents, along with their coreference chains.

Together, the results provide the first crosslingual picture of coreference preferences beyond the restricted entity-only patterns that the target of most existing work on coreference. The five languages are all shown to allow gradable use of personal and demonstrative pronouns for entity and event coreference, with systematic biases that align with existing generalizations about the use of lighter referential forms for less complex and more concrete referents.

5.3.2 Related Work

Referring expressions play a central role in human communication and are at the heart of multiple linguistics theories and applications involving natural language understanding. For both, a lot of attention has been put into the most standard cases represented by nominal expressions and personal pronouns while more atypical cases have been understudied. Among these is abstract anaphora – anaphora that involve reference to abstract entities such as events or states (Asher, 1993) whose complexity is increased because the form and interpretation of the referring expressions involved have a high degree of variability. This complexity is also behind the lack of resources annotated with this type of anaphora, simultaneously hindering its study.

5.3.2.1 Linguistic descriptions

Linguistic studies on the subject have employed both corpora and psycholinguistic methods. Corpus-based studies are a source of insights about language use, since the written texts they are based on are natural passages after all. They offer better estimates for building systems that will be used on those texts. On the other hand, corpus-based studies do not provide any explanation as to why a particular item follows a certain distribution, and they grant little control over the confounding variables responsible for that distribution.

In this respect, psycholinguistic studies are more suitable for capturing the cognitive processes behind naturally occurring phenomena. Psycholinguistic research has focused on using theoretical constructs of complexity, salience, and focus to capture coreference patterns.

In English, the anaphoric use of pronouns to refer to entities or abstract events has been evolving in the last centuries (Azuma, 2008): while *it* has consistently been the most used form to corefer to pronominal, highly accessible antecedents, the distal demonstrative *that* has grown in use to refer to NP or clausal antecedents (even overtaking the personal pronoun for non-nominal antecedents), positioning itself between *it* and the proximal demonstrative *this*, the most rarely used and marked of the three expressions. In an estimate of the current relative frequency of nominal coreference, Gundel et al. (2005) report that while about 84% of personal pronouns had nominal antecedents, only 28% of demonstratives did.

In other studies, the demonstratives *this* and *that* have been grouped together, assuming that they behave in the same manner and differently from *it*. Hedberg et al. (2007), for example, show how the two demonstratives tend to have non-nominal antecedents, unlike *it*: the personal pronoun requires the antecedent to be more highly activated, to be in the focus of attention, while demonstratives can retrieve antecedents that are less activated because of their complexity or because they were not directly introduced, such as events.

In a comparison between English, German, Italian and Spanish (Dipper et al., 2011), English and German were shown to prefer the use of demonstratives for retrieving abstract anaphors, while Italian and Spanish were not shown to display a preference. The German findings were further confirmed in a sentence rating study by Bentzen and Anderssen (2019), in which speakers of German were shown to prefer *das* to refer to non-nominal antecedents, with *es* being a less common but acceptable option when the antecedent is used as a "continuing topic". A corpus study comparing English, Danish and Italian (Navarretta, 2007) confirms this pattern, finding that in Italian the difference between personal and demonstrative pronouns with regard to their type of coreference is not significant,

unlike in English, where a contrast in use individuates different kinds of anaphors. In the Italian data, however, null pronouns were used for abstract referents in only 30% to 42% of instances (respectively in the original-Italian and parallel translations corpora). This is in keeping with antecedent accessibility hierarchies that position zero anaphora as requiring the highest level of accessibility (cf. Givón, 1983; Ariel, 1988, 2004). Assuming entities to be more stereotypical antecedents than events, as shown by crude annotation distributions (§5.3.1) the Italian and Spanish null pronouns should be employed less for abstract anaphora.

More generally, the possibility of referencing an event with *it* or *this* (or *that*) is captured in models of discourse deixis. In line with Bentzen and Anderssen's study, Webber (1991) posits that propositions, particularly those that are focussed, are referred to first with either *this* or *that* in immediately following sentences. In later instances of coreference with those same events, the pronoun *it* can then be used. Building on Centering Theory (Grosz et al., 1995), Passonneau (1989) analyzes intra-sentential instances of *it* vs *that* with an explicit NP antecedent, finding that *it* is used to refer to the center (most often the subject), whereas *that* favors the less prominent non-centers.

For French, Cornish (2015) shows how the demonstratives *cela* and *ça*, along with the neuter clitics *le*, *y* and *en*, can retrieve a sentential or verb-phrasal antecedent, and this relation is not simply formal-syntactic. A further claim is made comparing French with English, and saying that the anaphoric distinction between nominal and non-nominal antecedent has a correspondence to the forms: personal pronouns such as *she*/*he* and *elle*/*il* can only be used for nominal antecedents, while other expressions such as *do it*/*this*/*that* in English and *faire cela*/*ça* in French form a contrasting set used exclusively for non-nominal antecedents (such as predicates). Cornish goes on to observe that in French the nominative neuter pronoun *il* is rarely used as a propositional anaphor (which will use *ce*/*cela*/*ça*), because it is often used as an expletive.

In French corpus studies on the matter, Tutin (2002) shows the presence of demonstratives used as anaphors, although much less than personal pronouns and clitics, but the type of anaphora is not considered. Vieira et al. (2005), comparing French and Portuguese,

classify antecedents in the two categories of "concrete" and "abstract", where the first includes existing entities (such as people, things, places) and the second includes notions, actions and states. Their results show that, in both languages, demonstratives are used for abstract antecedents in more than three quarters of cases. However, their study targets demonstrative noun phrases, that is demonstrative adjectives rather than pronouns, so the results may not be comparable.

Here, we target these same previously studied languages, but we adopt a story continuation methodology to probe the availability of entity and event antecedents for demonstrative, personal and null pronouns. Where psycholinguistic approaches have been used previously to analyze *it* vs *that*, the emphasis has been on English. For example, Brown-Schmidt et al. (2005) report that comprehenders show a preference to interpret *that* to refer to a complex, composite antecedent (e.g., *I'll have the hamburger and fries. I'll have that, too.*), independent of other metrics of the salience of the referent. These results leave an open question regarding the crosslinguistic generalisability of such claims.

5.3.2.2 Computational approaches

Reference to non-nominal antecedents has largely been a niche area in computational linguistics research (see review by Kolhatkar et al., 2018). The most extensive annotation efforts in the field of coreference resolution have focused on English nominal coreference. OntoNotes (Pradhan et al.,2013), the largest and most frequently used corpus for training coreference resolution systems, for instance, only includes verbs if "they can be coreferenced with an existing noun phrase" according to its guidelines. Corpora with a richer annotation of event pronouns exist, but are much smaller. The most important resource is the ARRAU corpus (Uryupina et al., 2020), whose size amounts to about 20% of version 5 of OntoNotes. ParCorFull (Lapshinova-Koltunski et al., 2018) also contains annotations of event pronouns in English, German and French.

The scarcity of manually annotated resources has led to the use of artificial training data for coreference resolution systems of English non-nominal anaphora. Kolhatkar et al. (2013) study the resolution of anaphoric shell nouns such as "this issue" or "this fact" by exploiting cataphoric instances such as "the fact that...". Marasovic et al. (2017) construct

training examples based on specific patterns of verbs governing embedded sentences. Also on this front, Loáiciga et al. (2020) use parallel machine translation data in many languages for automatic data creation of event pronouns.

Before the breakthrough of neural end-to-end systems in coreference resolution (Lee et al., 2017), coreference resolvers needed to do explicit mention classification in order to exclude non-referential mentions before any resolution was attempted. In this context, the pronoun *it* has been targeted, as many of its uses are non-referential. Evans (2001) proposes the classification of the pronoun *it* into seven classes using contextual features. Boyd et al. (2005) report similar results of around 80% accuracy on the classification of instances of *it* using more complex syntactic patterns. Bergsma and Yarowsky (2011) describe a system for identifying non-referential pronouns using statistics of word sequences (*n*-grams) from the world wide web, however without accounting explicitly for event reference.

The many uses of *it* are also particularly relevant in dialogue, where event reference is much more common than in text data. However, there has been little recent research dedicated to coreference for general purpose dialogue systems, although current text coreference resolution systems are not trained to manage dialogue data. In this context, Müller (2007) proposes a disambiguation of *it* together with the deictic pronouns *this* and *that*. Both Eckert and Strube (2000) and Byron (2002) also reports on systems able to resolve both nominal and non-nominal anaphora. Working in a domain close to dialogue, Lee et al. (2016) create a corpus for *it*-disambiguation in question answering.

More recently, Loáiciga et al. (2017) proposed a semi-supervised setup based on a combination of syntactic and semantic features for the classification of *it*. Yaneva et al. (2018), on the other hand, report on experiments using features from eye gaze that prove to be more effective than any of the other types of features reported in previous works. This prior work points to the interest in understanding the behavior of pronouns like *it*, whose range of referential uses (entity, event, etc.) and non-referential uses presents a persistent challenge to computational systems. The approach we take here is to bring a technique from psycholinguistics to better model entity/event coreference.

5.3.3 Experiments

The work we present here tests entity/event coreference across five languages. We use a story continuation paradigm that presents a context sentence and then assesses how a subsequent referential form is interpreted. We collect story continuations in English, French, German, Italian and Spanish, measuring the rate of entity-versus-event coreference. We manipulate the referential forms (null versus personal pronoun versus demonstrative pronoun) along with two previously studied properties of the context sentence: aspect (perfective versus imperfective) and antecedent complexity (sentences with versus without an additional implicit argument, as determined by the alternating/non-alternating status of the verb). The goal is to use the same methodological probe and systematic context manipulations to test coreference biases across an inventory of coreferential forms in English, French, German, Italian and Spanish.

5.3.3.1 Choice of target pronouns

Our chosen target pronouns can be seen in Table 8. While the English pronouns are the same used in our previous study (Loáiciga et al., 2018), other factors guided the choice in the other four languages.

Table 8. Pronominal prompts in the five languages.

English	it, this
French	il, cela, c'est
German	es, das, dies
Italian	*null*, ciò, questo
Spanish	*null*, esto, este

Italian and Spanish allow a null subject, a case we decided to include in our study. Given the existence of theories pointing to a division of labour between null and overt pronouns in the two languages (see e.g. Carminati, 2002, for Italian, applied to Spanish in e.g. Alonso-Ovalle et al., 2002), and given hierarchies positioning null-anaphors at the highest point on accessibility/prototypicality scales (see Givón, 1983, and Ariel, 1988), it is interesting to check whether this also applies to event vs entity coreference: for example, a null pronoun may corefer more easily to an entity, while an event could be picked up by an overt form.

Other than the null, in Spanish we targeted the masculine proximal demonstrative adjective *este* and the neuter demonstrative pronoun *esto*. While *esto* is a singular-only pronoun, *este* can also be used as a determiner and its status is sometimes controversial (RAE, 2010: §17.2.2). Moreover, the Real Academia Española states that *esto* (and the other neuter demonstratives *eso* and *aquello*) are used to refer to inanimate entities or propositional content (while their use for animals is uncommon and for people, offensive) (§17.2.5b); however, the function of referring to propositional content is not mentioned for *este* (nor for the analogues *ese* and *aquel*).

In Italian we targeted the masculine proximal demonstratives *questo* and *ciò*: the two demonstratives were chosen following up from Navarretta's (2007) corpus study, in which *questo* had a slight preference for entities and *ciò* had a preference for events. Specifically, Serianni (1997: 198) describes *ciò* as a pronoun of neuter value (= this thing, that thing), and states it is of very common use, especially in written language, while spoken language would more commonly use *questo* (this) or *quello* (that).

For French, we chose the two pronouns *il* and *cela* and the expression *c'est*, composed of the particle *ce* and the copula in the third person singular of the indicative present tense. Our use of *c'est* reflected our concern that the bare form *ce* would be interpreted as a determiner (*ce chien*), limiting our ability to observe entity-vs-event coreference with the demonstrative NP *ce* itself. Using *c'est* prompts also allowed us to avoid the problem that the bare form *ce* would not have been compatible with a continuation with the verb *est* because of the obligatory vowel elision, while using the letter-punctuation combination *c'* in isolation posed a risk of confusion for the participants. *Il* is a third-person personal pronoun, it is the dummy subject in impersonal forms[22] and is used anaphorically to retrieve an infinitive or subjunctive clause. The demonstrative *cela* is a neutral form which is used for non-gendered antecedents like propositions. The Grevisse grammar (Grevisse and Goosse, 2008: §691ff.) includes in this neuter category, along with the distal pronoun *cela* and its proximal counterpart *ceci*, also *ça* and *ce*. *Cela* is more used in written language and can only refer to people in an informal register, while being normally used to refer to

[22] Interestingly, this function could historically be taken by *il* and *cela* alike, a use surviving in some expressions (e.g. *il/cela est vrai*, "it/that's true").

something "that we cannot name with precision" (§698.c). *Cela* and *ça* have replaced *ce* in most of its uses, but *ce* is still used for inanimate objects or to refer back to a sentence, and followed by a copula it refers to "what comes before or the situation" (§702).

For German, the Duden grammar (Dudenredaktion, 2009) says that *dies* is a shortened form of *dieses* which is predominantly used as a pronoun rather than a determiner, while *diese* can be used both ways (§372). In this account, phoric and deictic reference are distinguished such that the first links to referents without pointing explicitly while the second explicitly points to an object of discourse (§1818): textual content can be referred to phorically with *es* and deictically with *das* (§1821). Finally, while anaphoric personal pronouns can refer to nouns in distant sentences, the grammar explains that a demonstrative like *dieser* links anadeictically to the closest nominal candidate for reference (§1827), without addressing the potential of *dies* and *dieses* to refer to non-nominal antecedents.

5.3.3.2 Design and materials

The structure of the experiments in the five languages was the same. In each language, the 24 experimental items consisted of a sentence describing a situation followed, after a full stop and in the next line, by a pronominal prompt and a space to type in a continuation. For the null pronoun condition in Italian and Spanish, the continuation box was simply not introduced by a pronoun. The pronominal prompts used in the different languages are reported in Table 8, and an example of a prompt for each language is given in (13)–(17):

13) **English**: The colonial building was collapsing slowly. It/This ...
14) **French**: Le bâtiment colonial a croulé sous la neige. Il/Cela/C'est ...
15) **German**: Das prachtvolle Gebäude zerfiel über die Jahre. Es/Das/Dies ...
16) **Italian**: Il palazzo coloniale è collassato improvvisamente. Questo/Ciò/null ...
17) **Spanish**: El edificio colonial implosionó lentamente. Esto/Este/null ...

Participants of every group saw all 24 experimental items, presented with an event distribution of the language's prompts, interleaved with 42 fillers. Of these, 18 were items of an unrelated experiment involving named entities, 20 were real fillers including a context sentence concerning two referents followed by an adverbial prompt in a new sentence

(e.g., *However, Because of this*), and four were control items with unambiguous obvious responses (e.g., *Wilma played a Led Zeppelin guitar solo in front of the crowd. It was Stairway to* _____). The stimuli were as similar as possible across languages, deviating from a literal translation when it would not have sounded natural.

The stimuli also included manipulations of the status of the event as completed or ongoing. Half of the stimuli appeared in the perfective and half in the imperfective. The examples above exemplify a use of the imperfective aspect in English (13), and of the perfective aspect in the other languages (14–17). The tenses used were the present perfect and past continuous for English and their corresponding tenses in the other languages: the *passé composé* and the imperfect in French, the *passato prossimo* and imperfect in Italian, and the preterite and imperfect in Spanish. In German, where aspect is not encoded in verb tenses, the *präteritum* was used; the aspect could in this case be inferred either through adverbials or contextually.

Moreover, half of the stimuli had an alternating verb and half had a non-alternating verb (in a pattern that did not correspond to the perfective/imperfective split). The examples above all include a non-alternating verb.

Table 9. Summary statistics regarding the participants.

	N	Age: range	mean	σ
English	46	22–70	37.1	11.3
French	53	18–55	30.9	10.0
German	45	18–55	31.9	10.9
Italian	53	18–48	29.7	8.3
Spanish	51	18–67	33.0	9.7
	248	18–70	32.4	10.2

5.3.3.3 Participants and procedure

Participants were recruited on Amazon Mechanical Turk, targeting users from specific IP addresses (i.e. USA, France, Germany, Italy and Spain). They received $8/7€ for an estimated 45-60 minutes task. Seven of the German participants were excluded from analysis because their language competence (as judged by a native speaker) was obviously inadequate. Summary statistics for each group are reported in Table 9.

Continuations were collected via a web-based interface that participants could access from their own computer through MTurk. The website displayed a background questionnaire, a consent form and an instructions page, then each item was presented on a page by itself with a text box for participants to use for writing their continuation.

5.3.3.4 Annotation

The continuations were double-annotated for event or entity coreference. The annotators based their decision on the semantics of the continuation.

The pronominal prompt was hidden during the annotation process to avoid influencing the interpretation. Continuations which did not include a subject-position reference to the event or entity included in the prompt were excluded from analysis. Each example was annotated with a *strict* and a *liberal* interpretation. In the strict condition, an example was annotated as ambiguous if there was the slightest doubt about the correct reading. In the liberal condition, annotators were allowed to use their intuitive judgements when both readings were possible, but one of them seemed overwhelmingly likely. Both annotators were proficient speakers of the language and had experience relevant to language data.

To reconcile the double annotations, the following rules were applied:

- The labels used are: Event, Entity, Ambiguous, No;
- If both annotators agree in the strict annotation, the same label is used for the liberal annotation;
- When the two annotators disagree, the label No is used (either in the strict or the liberal annotation): these continuations, along with invalid or pleonastic ones, are then excluded from the analyses;
- In the liberal annotation scheme, if one annotator pins the reading down to either Event or Entity and the other labels it Ambiguous, the Event/Entity reading is chosen (i.e. if one annotator did not resolve an ambiguous reading, the opinion of the first annotator prevails): these continuations are only included in the liberal analysis.

5.3.3.5 Results with the strict annotation

The strict annotation was used as main analysis: comparing the same models computed with data annotated strictly and liberally, the models with strict data showed better fits (with, e.g., BICs being up to halved). The results from the liberal analyses are reported in §5.3.3.6 as a comparison with the strict analyses and in Appendix C.

All models were computed with the *lme4* package (Bates et al., 2015b) in R (R Development Core Team, 2008).

5.3.3.5.1 English

The analysis for English should replicate the results of Loáiciga et al. (2018), wherein we found a bias for *It* to corefer with entities and *This* with events, along with an effect of verb type, with verbs permitting alternation yielding more event coreference. The current analysis adds verbal aspect as a further predictor. The data was subset to continuations annotated as either event or entities (in the strict annotation scheme), leaving a total of 594 observations.

The selection of event vs entity coreference was modelled using generalised mixed-effects logistic regression (Bates et al., 2015b). The model predicted the isEvent binary outcome using condition (*it* vs *this*), aspect and alternation as binary predictors. All predictors were centred. To select the best fitting model, models with different fixed effects interactions were compared. The comparison showed that the interaction of verb alternation and aspect significantly improved the model fit over a model with no interaction ($p = 0.02$), while the interaction of condition and aspect or condition and alternation did not ($p = 0.74$ and $p = 0.07$, respectively). The three-way interaction did not improve the model's fit over a model with an interaction between aspect and alternation ($p = 0.78$). The chosen model thus includes fixed effects for condition, aspect, alternation, and the interaction between aspect and alternation.

The maximal random effect structure was used when supported by the data (Barr et al., 2013). Where a model did not converge, the random effects were successively removed, chosen by lowest variance. The maximal converging model includes random intercepts and slopes for condition by participant and by item. Following the recommendation of

Bates et al. (2015a), we ran a principal components analysis of the random effects structure. The PCA did not indicate any overspecification, so we selected the maximal converging model, whose results are reported in Table 10.

The model shows that Condition has a significant effect where the use of *This* increases event coreference ($p < 0.001$, see Figure 18). Verbs allowing alternation also increase event coreference ($p = 0.02$, see Figure 19). These results thus replicate those in Loáiciga et al. (2018). While aspect did not seem to have an effect on its own ($p = 0.97$), its interaction with verb type did: verbs that allow alternation yielded fewer event coreference in the perfective aspect ($p = 0.02$); essentially, using an alternating verb in the perfective rather than imperfective aspect inverts the general tendency of alternating verbs prompting more event coreference (e.g. from the data: *The last championship match started badly. It had a horrifying end too*).

5.3.3.5.2 French

The French analysis followed a similar procedure (cf. §5.3.3.5.1). The total number of observations was 741, and the three-way condition was coded as the differences "C'est – Cela" and "Il – C'est".

Model comparison showed that the best model was one with no interactions, with e.g. the addition of all interactions not significantly improving over no interactions ($p = 0.4$). The only converging random effect structure was the minimal one, with random intercept by participant. The results of the model are reported in Table 10.

The fixed effects show a significant difference between *Il* and *C'est* ($p < 0.001$), with *Il* yielding fewer event continuations than either of the other two variants. The difference between *C'est* and *Cela*, on the other hand, is not significant ($p = 0.49$). Alternating verbs, as in the other languages in which it is significant, prompt more event coreference than non-alternating verbs ($p < 0.001$, see Figure 19), while the effect of aspect did not reach significance ($p = 0.25$).

Given the three-way nature of condition, its effects were confirmed subsetting the data to two of the three variants. In the model with *Il* and *C'est* only, as well as in the that with *Il*

and *Cela* only, condition proved significant (both $p < 0.001$ and with the same direction of effect as in the full model); in the model with *C'est* and *Cela* only, condition was not significant ($p = 0.948$), confirming that *C'est* and *Cela* do not pattern significantly differently in how they bias event or entity coreference (cf. Figure 18). In all three, alternation was significant (always $p < 0.01$ or $p < 0.001$).

5.3.3.5.3 German

The German data was checked by a native speaker and subset to exclude seven participants whose language competence did not seem native. A further subset only included data "strictly" annotated as either event or entity, leaving a total of 612 observations. All factors were centred: as condition is a three-way factor (*Es* vs *Das* vs *Dies*), it was centred as the difference between pairs of its values, i.e. "Dies – Das" and "Es – Dies".

Following a similar procedure as that outlined in §5.3.3.5.1 and §5.3.3.5.2, models were compared to select significant interactions. However, no interaction significantly improved the model fit (e.g. the model with s interactions' fit improvement significance over the null was $p = 1$), so the chosen model includes predictors for the condition, aspect, and alternation with no interactions.

Again, we started from a maximal random effect structure and reduced it until model convergence, after which we ran a principal components analysis of the random effects structure which did not point to any overspecification. The maximally converging model includes random intercept and slope for Condition by participant, and a random intercept and slope for Alternation by item. The results of the model are reported in Table 10.

The model shows a general bias towards events rather than entities ($p = 0.001$). There is also a significant difference between *Es* and *Dies*, wherein *Dies* too yields more event coreference than *Es* ($p < 0.001$). *Dies* and *Das* are not significantly different in their influence on coreference ($p = 0.69$, cf. Figure 18). Alternation was significant in that alternating verbs increased the chance of event coreference ($p = 0.03$, cf. Figure 19), but no significant effect was found for Aspect ($p = 0.78$).

The effects of condition were confirmed subsetting the data into couples of two of the three possible pronouns and fitting new models. These three models use minimal random effect structure as it was the only common converging structure. In the subset with *Es* and *Dies* only and in that with *Das* and *Es* only, condition was significant ($p < 0.001$), while in the subset with *Dies* and *Das* only it was not ($p = 0.69$), confirming that *Dies* and *Das* do not have different effects on the choice of coreference (cf. Figure 18). Alternation was only significant in the model with *Es* and *Dies* only ($p = 0.04$) and, surprisingly, aspect was significant in the model with *Dies* and *Das* only ($p = 0.04$), with perfective verbs yielding more event coreference.

5.3.3.5.4 Italian

The Italian analysis also followed a similar procedure (cf. §5.3.3.5.1-§5.3.3.5.3). The three-way condition was coded as the differences "Ciò – Null" and "Questo – Ciò". The total number of observations was 413.

Model comparison showed that the interaction of Condition and verb Alternation significantly improved the model fit over a model with no interactions ($p = 0.01$), while the improvement of the maximal model with all interactions did not outweigh the increased complexity ($p = 0.34$). To reach model convergence, the random effects only included random intercepts by participant and item. The model results are reported in Table 10.

The model output shows a significant difference between the null pronoun and *Ciò* ($p < 0.001$, see Figure 18), whereby the null pronoun yields fewer event continuations, while the difference between the two overt pronouns is not significant ($p = 0.95$). The intercept shows a strong general tendency towards events ($p < 0.001$). Neither verb Alternation nor Aspect reached significance (respectively, $p = 0.13$ and $p = 0.15$). No interaction reached significance.

Again, the effects of condition were confirmed subsetting the data to two of the three variants. These models only included an intercept by participant as random effect. Condition proved significant in the model with the null pronoun and *Ciò* and in that with the null pronoun and *Questo* (both $p < 0.001$), but not in the model with the two overt pronouns (p

= 0.79). The direction of the effects is the same, confirming that the pronoun with a different pattern with regard to event vs entity coreference is the null pronoun, which is biased towards entity continuations. Alternation was significant with the usual direction in the model with *Ciò* and *Questo* (p = 0.04), and its interaction with Condition was significant in the model with the null pronoun and *Questo* (p = 0.01), with null pronouns being used proportionally less for events with alternating verbs.

5.3.3.5.5 Spanish

The Spanish analysis followed a similar procedure as the other languages (cf. §5.3.3.5.1-§5.3.3.5.4). The three-way condition was coded as the differences "Esto – Este" and "Null – Esto". The number of observations was 505.

Through model comparison, we chose the model including the three predictors but none of the interactions (as they did not significantly improve the model fit, e.g. all interactions does not improve over no interactions: p = 0.051), while the maximally converging random effect structure included random intercepts by participant and item. The results of the model are reported in Table 10.

The model shows a significant difference between *Esto* and *Este*, and between the null pronoun and *Esto* (both p < 0.001): *Esto* yields more event coreference than either *Este* or the null pronoun. While Aspect did not seem to have an effect on event vs entity coreference (p = 0.15), verbs allowing Alternation were biased towards event coreference (p < 0.001).

The effects of condition were confirmed subsetting the data to two of the three pronouns. Condition was significant in all models: in the model with the null pronoun and *Esto* only, as well as in the that with *Este* and *Esto* only, with the same direction of effect as in the full model (both p < 0.001); in the model with the null pronoun and *Este*, with *Este* yielding fewer events than the null pronoun (p = 0.004) (cf. Figure 18). Alternation was significant in all three models (with significance between p = 0.02 and p = 0.001), and Aspect was significant in the model with only the null pronoun and *Esto* (p = 0.03), with perfective verbs yielding more event coreference.

Table 10. Estimated models fixed effects (strict annotations).

Effect	Estimate	Std.Error	z value	$Pr(>\|z\|)$	
			English		
(Intercept)	0.10	0.29	0.35	0.72	
Condition (this)	5.22	0.63	8.27	< 0.001	***
Aspect (perf)	-0.01	0.31	-0.04	0.97	
Alternation (alt)	0.75	0.33	2.30	0.02	*
Aspect:Alternation	-1.59	0.66	-2.39	0.02	*
			French		
(Intercept)	0.01	0.14	0.07	0.95	
C'est − Cela	-0.16	0.23	-0.70	0.49	
Il − C'est	-4.38	0.35	-12.38	< 0.001	***
Aspect (perf)	0.25	0.22	1.16	0.25	
Alternation (alt)	0.94	0.22	4.22	< 0.001	***
			Spanish		
(Intercept)	0.50	0.28	1.82	0.07	
Esto − Este	6.56	0.74	8.83	< 0.001	***
Null − Esto	-5.22	0.73	-7.15	< 0.001	***
Aspect (perf)	0.56	0.39	1.42	0.15	
Alternation (alt)	1.58	0.43	3.66	< 0.001	***
			German		
(Intercept)	3.05	0.94	3.24	0.001	**
Dies − Das	-1.05	2.62	-0.40	0.69	
Es − Dies	-8.61	1.76	-4.88	< 0.001	***
Aspect (perf)	0.78	0.42	1.85	0.06	
Alternation (alt)	0.94	0.44	2.13	0.03	*
			Italian		
(Intercept)	1.99	0.41	4.81	< 0.001	***
Ciò − Null	5.78	0.97	5.96	< 0.001	***
Questo − Ciò	-0.05	0.71	-0.07	0.95	
Aspect (alt)	0.77	0.53	1.44	0.15	
Alternation (perf)	0.66	0.44	1.50	0.13	
(Ciò − Null):Altern	0.95	1.11	0.86	0.39	
(Questo − Ciò):Altern	2.44	1.49	1.64	0.10	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

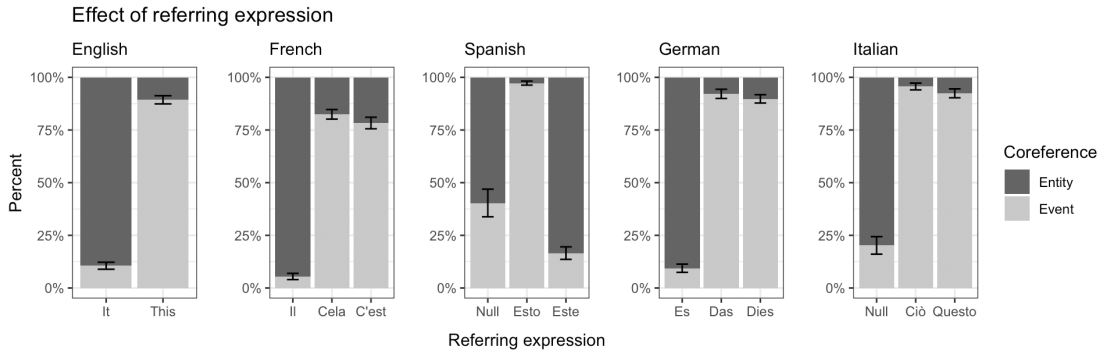

Figure 18. Event and entity coreference by pronoun prompt in the five languages.

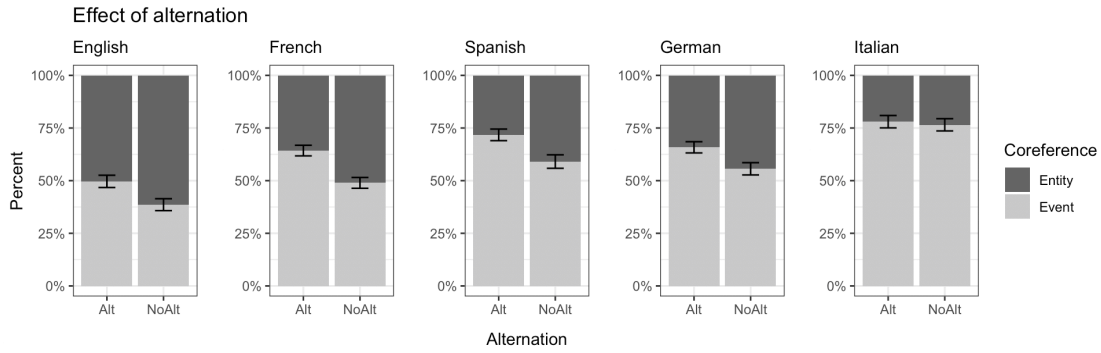

Figure 19. Event and entity coreference by verb type in the five languages.

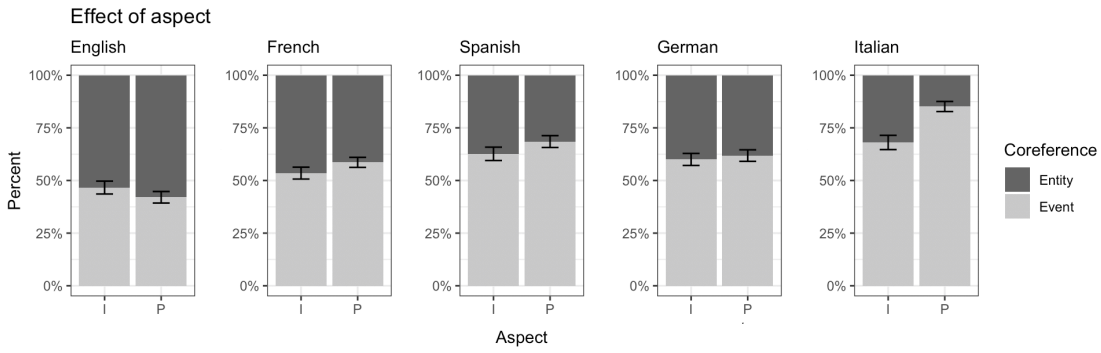

Figure 20. Event and entity coreference by verb aspect in the five languages.

5.3.3.6 Results with the liberal annotation

As previously mentioned, the models using the liberal annotation of event or entity score worse in measures of the relative model quality like the BIC, AIC and log-likelihood. We therefore chose to base our analysis on the models with strict annotation. Although the results of the two types of annotation are very similar, it is worth pointing out some of the differences. The full outputs of the models are reported in Appendix C.

One of the effects that did not significantly improve the models in the strict annotation did with liberally annotated data. Specifically, in French the interaction between Aspect and the Referring Expression condition improved the model. One of these two interactions reached significance, namely that when a perfective verb is used, *Il* yields even fewer event continuations than *C'est* (p = 0.05).

In German, the liberal model did not show any significance in the effect of alternation (p = 0.52, with same direction of the effect). By contrast, in the Italian liberal model the effect of Aspect had a significance which is not reached in the strict model (p = 0.15, same direction of the effect).

The models for English and Spanish yielded the same results in both annotation guidelines.

It is worth noting that the effect of Aspect, which almost disappears across the board in the strict modelling, in some cases reaches significance in the models subsetting the data in three-way factors, as in German (§5.3.3.5.3) and Spanish (§5.3.3.5.5).

5.3.4 Discussion

The results of the five studies display similarities between our target languages.

In all languages, heavier referring expressions (specifically, demonstratives) bias coreference towards event continuations (as shown in Figure 18). This is in keeping with theories positing that richer, more uniquely-referential expressions (like demonstratives, that add a distal trait to the those of personal pronouns and null forms) will be used to retrieve less "stereotypical" material (e.g. Ariel, 1988): given that event coreference is much rarer than entity coreference (cf. §5.3.1), it is likely that it will be implicitly treated as a marked case,

granting the use of expressions used for less accessible antecedents. A notable exception to this pattern is the Spanish demonstrative *este*, which was not only biased towards entity continuations, but even more so than the null pronoun. One possibility is that, unlike its neuter counterpart *esto*, the masculine form *este* could be used preferably for (masculine) entities, for analogy with the feminine form *esta*. Note that the masculine is the default gender of the entities used in the experiment.

The results of the studies often agree with the (rough) prescriptions of grammars. The RAE (2010) predicted the use of Spanish *esto* to refer to things other than entities, and similarly in French the Grevisse and Goosse (2008) grammar predicted the use of *cela* and *ce* for non-entities. The Italian data agrees with Serianni (1997): both *ciò* and *questo* are dispreferred for entity antecedents. For German, the Duden (2009) grammar does not provide a very clear indication of use, in that it says that both *es* and *das* can be used to refer to non-entities (however differently) – yet, our data shows a marked tendency for *es* to refer to entities. However, our studies differ from these prescriptions in that they offer an estimation of the degree to which the different forms are non-categorical.

Comparing the results with those of corpus and psycholinguistic studies also yields some interesting observations. The English results confirm a contrast in use between the personal and demonstrative pronouns already found in Loáiciga et al. (2018), Navarretta (2007) and Dipper et al. (2011), which also agrees with our results on German, in which *es* continues referential chains of topical entities while *das* refers to non-nominal antecedents. Dipper et al. (2011) does not find a clear distinction in use between demonstrative and personal pronouns in Spanish, and our results add some detail to the claim showing how *este* does not pattern as expected. In Italian, Navarretta (2007) finds that while *ciò* is used for non-entities, *questo* has a light preference for entities: this is in contrast with the very clear tendency we found for the demonstrative to retrieve events. The French results are in line with Vieira et al. (2005) in that demonstratives are used for abstract antecedents more often.

Finally, our studies targeted two verbal features: alternation and aspect. Alternating verbs yielded more event continuations in more languages, even if the effect did not reach significance in Italian. In Italian, the two demonstratives show ceiling selection of the events as antecedents: this near-categorical behaviour may have obscured the effect. The cross-linguistic effect of alternation is visualised in Figure 19.

On the other hand, aspect did not show a significant effect across languages; nonetheless, the direction is quite consistently the same in which the representation of an event as completed through a perfective aspect yields more events (except for English: see Figure 20). This, however, was only significant in the Italian model run with the liberal annotation of our data and in interaction with alternation in English, whereby non-alternating verbs yield more event continuations in the perfective aspect, whereas alternating verbs do so in the imperfective.

5.3.5 Comparison with annotated coreference

In order to contexualise our findings in the coreference resolution scenario, we take a multilingual parallel corpus to check whether the same patterns observed in the previous experiments surface from the annotated data.

We work with the ParCorFull corpus, which includes annotations for the English, German and French languages. The underlying coreference scheme was designed for uniform coreference annotations across the languages (see Lapshinova-Koltunski and Hardmeier, 2018 for details). The annotated elements (markables) in this corpus include pronouns, nouns, nominal phrases or elliptical constructions that are parts of a coreference pair (antecedent-anaphora), as well as verb phrases or clauses being antecedents of event anaphora. The annotated antecedents are of two different types: entities and events. Entities can be either pronouns or noun phrases, whereas events include verb phrases, clauses, or a set of clauses. ParCorFull includes texts from TED talks transcripts and newswire data, and comprises approximately 160,000 tokens.

5.3.5.1 Event vs entity antecedent proportions

To reproduce the parameters of the experiments described above, we first extract all coreference chains headed by lexical entity or event antecedents, thus excluding cataphora.

We then retain all mentions of the pronouns of interest (EN: *this*, *it*; DE: *es*, *das*, *dies*; FR: *c'*, *il*, *cela*) in a subject position and order them according to their appearance in the text.

Table 11 summarises the distribution of the mentions retained after filtering. The index i = *n* reflects the order of re-mention of the antecedent by a pronoun of interest. *i* = 1 corresponds to equivalent cases to those produced in the experiments with human participants, where a particular type of antecedent is re-mentioned for the first time with one of the different pronouns. *i* = 2 corresponds to the second re-mention of an antecedent, *i* = 3 to the third, and so forth. Note that while this is a parallel corpus, the number of times there is a re-mention of an antecedent varies in each of the languages, with French being the language with the most re-mentions. Additionally, for all the languages, it can be seen that there is a clearly preferred form for re-mention consistently used through time as the text increases in length, regardless of the type of antecedent. These are light forms: *it* for English, *es* for German, and *il* for French.

This pattern seems coherent with Webber's (1991) proposal that demonstrative pronouns are used to refer for the first time to parts of the context which are in focus, making them available to serve as antecedents for personal pronouns later. Moreover, it shows how demonstratives give rise to a topic shift that enhances the prominence and subsequent persistence of a referent (Givón, 1983): an inverse relation is then shown whereby higher prominence leads to longer chains of re-mention in the discourse (von Heusinger and Schumacher, 2019). While in a first re-mention a referent may not be an established discourse topic yet, it will become more and more ratified with further re-mentions, thus granting the use of lighter referring expressions.

A direct comparison between the experiments with human participants and the corpus data is summarised in Table 12. As it is observed in the human continuations, events are mostly referred by demonstrative pronouns while entities by personal pronouns, but this preference is not categorical. In English, it is particularly interesting that events in the corpus have a similar proportion of *it* vs *this*. Something similar happens with French *c'* vs *cela*, while in German the clear preference is *das*. Another interesting figure is that while the human responses are perfectly balanced, as a result of controlling the parameters for the

experiment, the proportion of events in the corpus is around 35% in contrast to 65% entities.

Table 11. Re-mention frequency of entity and event antecedents in English, German and French as annotated in the ParCorFull corpus. The index *i* = *n* represents the subsequent order in which the pronoun appears after an antecedent has been introduced.

Antecedent	Anaphor	Re-mention index											
English		$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	$i=7$	$i=8$				
Entity	*this*	40	8	0	1	0	1	0	0				
	it	284	87	35	16	12	5	1	1				
Event	*this*	68	5	3	0	0	1	1	0				
	it	91	33	5	5	1	0	0	0				
German		$i=1$	$i=2$	$i=3$	$i=4$	$i=5$							
Entity	*es*	108	46	16	6	3							
	das	130	10	2	1	0							
	dies	10	1	0	0	0							
Event	*es*	25	14	5	2	1							
	das	178	20	2	0	0							
	dies	9	2	0	1	0							
French		$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	$i=7$	$i=8$	$i=9$	$i=10$	$i=11$	$i=12$
Entity	*c'*	51	7	2	0	0	0	0	0	1	0	0	0
	il	70	23	10	10	7	6	4	4	2	3	1	1
	cela	4	0	0	0	0	0	0	0	0	0	0	0
Event	*c'*	58	3	1	2	0	0	0	0	0	0	0	0
	il	7	5	1	1	2	0	0	0	0	0	0	0
	cela	22	0	2	0	0	0	0	0	0	0	0	0

Table 12. Comparison between the proportions of event and entity antecedent interpretations in the human experiments with the annotations in ParCorFull of antecedents and their first re-mention (*i* = 1). Note that the percentages are computed using the total counts per language.

	Antecedent	English		German			French		
		this	*it*	*es*	*das*	*dies*	*c'*	*il*	*cela*
Human responses	Entity	27 5%	305 51%	203 33%	12 2%	24 4%	48 6%	225 30%	48 6%
	Event	226 38%	36 6%	21 4%	141 23%	211 35%	117 24%	13 2%	230 32%
Corpus annotation	Entity	40 8%	284 59%	108 23%	130 28%	10 2%	51 24%	70 34%	4 2%
	Event	91 19%	68 14%	25 6%	178 39%	9 2%	58 27%	7 3%	22 10%

5.3.5.2 Governing verb alternation status

In order to draw a similar comparison with respect to the alternation status of the verbs in the corpus, we also extracted the verbs relevant for the entities and events reported in Table 12. In the case of entity antecedents, we extracted the verb to which the head of the noun phrase is attached. For event antecedents, we extracted the verb from the antecedent itself.

Studies about the causative alternation propose that verbs can be ranked on a universal scale of likelihood of spontaneous occurrence (Haspelmath, 1993). In this scale, verbs are ranked according to the degree to which they are non-agentive (Samardžić, 2014, p. 180). This means that verbs ranking higher in spontaneous occurrence can occur without an explicit agent causing the event, e.g., *break*, *open*, and are hence more likely to participate in the causative alternation. Following Haspelmath and using large scale corpus data, as opposed to typological observations, Samardžić estimates the degree of spontaneity for the 354 English verbs permitting the causative alternation reported by Levin (1993). The underlying assumption for the spontaneity score is that causative and anticausative uses of a verb have a correlation with transitive and intransitive examples in corpora (reported to be $r = 0.67$, $p < 0.01$ using the Spearman test).

We then compare the verbs extracted from the ParCorFull antecedents against the list of scored verbs reported in Samardžić (2014). Unfortunately, being restricted to Levin's verbs, this list is very small, and contains only a few of our verbs. The extracted English verbs which matched the list are listed in Appendix D. We could not find a similar resource for French or German.

It is very difficult to estimate a similar spontaneity score for our verbs since, unlike Samardžić who worked with Levin's list and the Proposition Bank (Palmer et al., 2005) which is annotated for semantic roles, we have many verbs for which we do not have any sort of gold standard resource that help us estimate the accuracy of our calculations. In our case, equating transitive uses with alternating verbs (causatives) and intransitive uses with non-alternating verbs (anti-causatives) is a dangerous simplification since many verbs in

the corpus do not necessarily have the proper frame of thematic roles playing a part in the causative alternation.

Interpreting and generating coreference involves many levels of linguistic processing, including verb semantics (see e.g. Arnold, 2001, for the effect of thematic roles). Between our experimental items, we manipulated the alternation status of the verb because the experimental framework grants us complete control over the stimuli. Estimating semantic predictors from a corpus is much harder, partly because of the lack of resources annotated with this type of information. With our work, we highlight the value in considering semantic predictors for the processing of coreference.

5.3.6 Conclusions

Our study shows a crosslinguistic bias whereby the coreference to entities or events is biased by multiple factors: the referring expressions used and features of the verb constituting the argument structure involving the possible entity antecedents. Moreover, our data gives a clear distribution of event and entity coreference across referring expressions and languages, with strictly controlled items achieving less noise than in a corpus analysis.

Confirming the predictions based on hierarchies describing the use of referring expressions based on their antecedent's accessibility (e.g. Ariel, 1988), our results show that lighter referring expressions are biased towards entity antecedents and, conversely, heavier expressions are biased towards events. The lesser accessibility of events can be explained in multiple ways: they are a less common type of antecedent (as corpus measures show: see §5.3.1, §5.3.5), they are more complex and less easily introduced directly (Hedberg et al., 2007).

The comparisons with a coreference-annotated corpus generally replicate the human results, but they also suggest that these patterns might be further influenced by the time at which an antecedent is re-mentioned in a coreference chain. The corpus results also show slight differences in the referring expressions' proportions, which may be due to stylistic variation. In addition, the fact that the different forms appear non-categorically with either entities or events points towards an interplay between both types of antecedents in

the discourse, indicating that studying one without the other might result in an incomplete picture.

We also investigated verbal features and their possible influence on coreference. On the question of verb aspect, this study saw a general tendency for completed events to be taken up as an antecedent more than ongoing events in all languages but English, but this effect did not reach statistical significance. Further research with fewer confounding factors is needed to confirm whether the mostly uniform direction it showed in our data is upheld.

On the other hand, the event structure of verbs was also shown to clearly influence the coreference patterns. Specifically, further increasing the complexity of an event by having an implicit agent in alternating verbs creates more competition for the explicit entity: since the accessibility of a referent decreases with the increase in the total number of possible referents (Grosz et al., 1995), the event antecedent becomes relatively more accessible and is chosen more as a likely continuation.

The crosslingual effect of verb semantics shows that, from a cognitive point of view, adding a competitor to a pool of possible antecedents affects the relative accessibility of the other antecedents even when said competitor is not made explicit. This raises questions about whether introducing implicit and explicit elements has the same effect on coreference (both qualitatively and, especially, quantitatively), and whether an antecedent's conceptual availability as given by world knowledge matters as much as its explicit presence in a discourse. Such differences in the activation of antecedents could be explored via "online" psycholinguistic methods (e.g. measurement of reaction times or eye-tracking). We must leave these questions for future research.

5.4 Conclusions

In the current chapter, two studies were reported on the coreference of events versus entities. Story continuation experiments in five languages biased the coreference choices of participants manipulating anaphoric referring expressions, to test whether light versus heavy referring expressions (personal and demonstrative pronouns) had a specialised use

in retrieving simple versus more complex and rare antecedents (entities and events respectively). Verb features were furthermore manipulated to represent events as either ongoing or completed, and to examine whether the possibility some verbs have to appear in both causative and inchoative constructions (thus making an implicit agent available even in inchoative frames like *the snow melted*) affects the prominence and retrievability of the antecedents (as predicted by Centering: Grosz et al., 1995).

The results of the studies showed a crosslinguistic tendency whereby personal pronouns favour entity antecedents and demonstratives favour event antecedents (as predicted by e.g. Accessibility Theory: Ariel, 1988; and replicating Çokal et al., 2018). The first study (§5.2) also showed the general intention of the speaker/writer in the first sentence to affect coreference: using stimuli extracted from a corpus, the original mention of an entity or an event in the corpus affected participants' choices in the experiment. This was examined in more detail, and crosslinguistically, in the second study (§5.3), finding that both verb class and aspect (although less consistently) bias coreferential choices, with the availability of an implicit entity increasing event coreference (as seen for English in §5.2), and perfectivity also increasing event coreference.

These results point to entities and events being affected in different ways by prominence, with their levels of prominence influenced by minute changes in the sentence such as type of verb and the number of its arguments, as reflected in the interpretation of different types of referring expressions. Moving from the sentence level to the entities itself, the next chapter (6) will examine what affects the prominence not of different components of an event, but of the components of a named entity: its grammatical number (i.e. the interpretation of a collective noun as a single unit) or its notional, semantic number (i.e. its interpretation as a group of multiple smaller elements).

Chapter 6 Named entity conceptualisation: the whole versus its parts

What components of an entity are more prominent?

6.1 Overview

The last chapter of this thesis focusses on named entities, in an attempt to look at the relative prominence of different components of meaning within a single referent. A named entity is a type of collective noun, and as such they show a conflict between grammatical and notional number: the noun *collective*, for example, is grammatically singular, but semantically it may be considered a plural because it is clearly composed of many parts, the members of the collective. However, named entities have proper names (denoting a brand, company, music band, sports team, etc.) which are not marked for number, yielding ambiguity between the two types of number.

When referring to a named entity, a speaker will have to make a decision with regard to number agreement on the verb and in anaphoric expressions. Either the grammatical or the notional number will be seen as more prominent, and based on how the entity is indexed, speakers will refer to it as a whole or as the components it comprises. This number conceptualisation in anaphoric reference has been shown to be affected by the abstractness of the antecedent, with more concrete entities in which the parts are more clearly individuated leading to more plural concordance and less concrete entities leading to more singular concordance (Eberhard 1999). Moreover, the choice of grammatical number may change in time and among different varieties (Hundt, 2006; 2009), or even diaphasically, with formal genres more strictly adhering to prescriptive rules (Levin, 2001; Hundt, 2009).

Moreover, the choice of number for the referring expression may conflict with other choices, such as that of gender in languages where it is marked. Grammatical gender, like

number, is not morphologically explicit in the proper names of most named entities. Different referring expressions either force a choice on the speaker, or allow them to avoid the choice altogether, as shown in the German examples in (1):[23]

1) Google trat 1998 in den Suchmaschinenmarkt ein. Zehn Jahre später
   Google entered the search engine market in 1998. Ten years later
   a. ...ist er Marktführer. / ...war es der Marktführer.
      ...he is the market leader. / ...it was the market leader.
   b. ...sie sind die Marktführer in den Suchmaschinenmarkt.
      ...they are the market leaders in the search engine market.
   c. ...war Google bereits Marktführer.
      ...Google was already the market leader.

While in (1a) the speakers conceptualise "Google" in the singular and commits to a gender with the personal pronoun (respectively, with the masculine pronoun *er* and the neuter *es*), in (1b) the plural conceptualisation avoids a choice on gender (with the plural pronoun *sie*, which is not gender marked). Finally, in (1c) the repetition of the proper name also avoids a gender commitment, while agreeing by number in the singular (plural agreement would have also been possible, although not in all cases and languages; cf. §6.3.3).

The choice to corefer to a named entity with its grammatical or notional number may vary for a range of reasons: other than crosslinguistically, following from the different typological traits of languages, the selection of number may change with different types of entities and with the genre of text, especially when precise prescriptive rules exist within a language. To look at these factors, this chapter presents two studies, one on English (§6.2) and one comparing five languages: English, French, German, Italian and Spanish (§6.3). The first study employs a corpus analysis targeting genres at different levels of formality and two story continuation experiments, with either corpus-extracted or constructed stimuli. The results show a strong preference in English for plural conceptualisation with

[23] The prompt sentence and natural continuations are taken from the data from the study presented in §6.3.

the pronoun *they* as a referring expression for the informal user-generated story continuations, and an effect of formality whereby more formal genres employ the singular conceptualisation more, using either the pronoun *it* or a repetition of the entity's name.

The second study employs a story continuation paradigm to study the number conceptualisation of named entities crosslinguistically, examining whether systemic properties of each language, such as the grammatical gender marking and the availability of non-gender-marked referring expressions, are influential in speakers' choice of referring expression. Referring expressions that are not marked by gender, such as the repetition of the entity's proper name, may violate information structural preferences such as the avoidance of semantically heavy forms to refer to topical antecedents (e.g. the Repeated Name Penalty; Gordon et al., 1993). The results show that gender-avoiding strategies are consistently used, with Italian and Spanish using null pronouns and French and German using more names and nouns despite the information structural penalties associated with them. Italian and Spanish also exhibit the most singular conceptualisation (grammatical agreement), with German and English using the most plural continuations (notional agreement). Finally, more imageable entities, made of more clearly individuated parts, were shown to be more likely to be conceptualised as a plurality.

Both studies were conducted in collaboration with the same research group as the studies in Chapter 5, namely Hannah Rohde, Sharid Loáiciga, Christian Hardmeier, and by myself. The first article (§6.2) was included in the *Seventh Named Entities Workshop (NEWS)*; the second article is yet to be submitted to a journal. In both studies, I took part in the design, testing and analysis of the psycholinguistic components, while I was involved only at a higher level in the corpus studies. While I am principally responsible for the writing of the second article in the version presented in §6.3, the first article was written collectively.

6.2 Forms of anaphoric reference to organisational named entities: Hoping to widen appeal, they diversified[24]

6.2.1 Introduction

The names of organisations such as political bodies or companies are often made-up words (e.g. "Intel", "Novartis") or acronyms (e.g., "EU", "Unesco"). They differ from other noun phrases in that they offer very little information about their grammatical properties such as number or, in languages where this is relevant, gender. Such names are a special case of the broader category of collective nouns, which also includes common nouns such as "team" or "committee", and they can be conceptualised in different ways by focusing on the collective as a singular unit or on the plurality of people which the organisation is comprised of. When they occur as antecedents of referring expressions, names of organisations are a challenge for natural language processing (NLP) because they can trigger different types of morphological marking on the anaphoric elements. Moreover, the preference for certain types of agreement varies across different genres and, we expect, different languages. The experiments presented here address English only and serve as a pilot study for an investigation of reference to organisations across multiple languages.

Via a corpus analysis of the OntoNotes corpus (Pradhan et al., 2013) and two crowdsourced story continuation experiments, we study how organisational named entities are referenced after their introduction in a discourse. Specifically, we consider anaphoric expressions coreferent with the proper name of an organisation that are separated from their antecedent by a sentence boundary, but no intervening mentions belonging to the same coreference chain. The expressions are categorised into four classes: repetition of the proper name (name), paraphrastic noun phrases with a common noun such as "the company" (noun), and forms of the pronouns *it* and *they*. The pronominal case is informative to speakers' choice between a conceptualisation as singular (*it*) or plural (*they*).

[24] Hardmeier, C., Bevacqua, L., Loáiciga, S., & Rohde, H. (2018). Forms of anaphoric reference to organisational named entities: Hoping to widen appeal, they diversified. In *Proceedings of the seventh named entities workshop*, 36–40. Association for Computational Linguistics. Available at https://www.aclweb.org/anthology/W18-2406.

6.2.2 Related literature

Morphological agreement with collective nouns has received some attention in English linguistics, but most research focuses on the agreement of verbs rather than pronouns, and – to an even larger extent – on collective common nouns such as "team", which are formally singular but can trigger plural agreement, rather than proper names.

There is broad agreement that American English prefers singular verb agreement with collective nouns, whereas notional concord with plural forms is not uncommon in British English (Fries, 1988; Bock et al., 2006; Hundt, 2009). Other varieties of English range in between (Hundt, 2006). Shift towards singular agreement is considered to be an ongoing diachronic process (Hundt, 2009), but the extent to which plural verb agreement with collectives is disappearing among younger speakers of British English is disputed (Fries, 1988).

The situation for pronouns is different. Pronouns following collective nouns are more likely to receive plural marking than verbs (Hundt, 2006,2009), particularly in speech (Levin, 2001), and there is psycholinguistic evidence of processing differences favouring syntactic (singular) agreement for verbs and notional (plural) concord for pronouns (Bock et al., 2006). Singular and plural agreement can also co-occur with the same mention ("mixed concord"), typically involving a singular verb and a plural pronoun (Hundt, 2009).

6.2.3 Corpus analysis

6.2.3.1 Corpus and extraction

The OntoNotes corpus (Pradhan et al., 2013) contains about 1.7 million words of annotated English text predominantly of American origin from different genres, or data sources: *newswire* (nw), *broadcast news* (bn), *broadcast conversation* (bc), *magazine* (mz), *telephone conversation* (tc), *webdata* (wb) and *pivot text* (pt).[25] We extract examples using the gold-standard annotations of coreference and named entity type. Each example is a pair of mentions belonging to the same coreference chain. To ensure that the corpus analysis is comparable with the continuation studies described in Section §6.2.5, we only extract pairs of mentions in adjacent sentences, excluding both pairs of mentions in the same sentence and

[25] The pt subcorpus contains excerpts of the Bible and is not used in this paper.

pairs with intervening sentences. A pair of mentions is extracted if the two mentions are neighbouring members of the same coreference chain (i.e., no mentions of the same chain occur in between) and the first mention is annotated as a named entity of type ORG.

6.2.3.2 Overview

Table 13 and the first six bars of Figure 21 show the distribution of reference types for the different OntoNotes genres. The size of the individual sub-corpora varies substantially and so does the number of examples that can be extracted from each. The smallest non-empty sample ($N = 12$) is from the *telephone conversations* (tc) subcorpus, the largest ($N = 1242$) is from *Newswire*.

The most common type of reference, making up 58–75% of the examples in all subcorpora, is a repetition of the name. Paraphrasing noun phrases are more common in *broadcast news* (19.6%), *magazine* (16.3%) and *newswire* (15.5%) than in *web data* (10.8%) and *broadcast conversation* (9.5%). Many examples in the *other* category are instances of the first-person pronoun *we* that occur when a representative of the organisation is quoted or speaking. The relative frequency of pronominal references (*it* and *they*) varies considerably between genres. It is greatest in *telephone conversations*, where 5 out of 12 references are of this type. In *newswire* (7%) and *broadcast news* (10.2%), pronominal references are much less common. *Web data* (13.5%), *magazine* (19.0%) and *broadcast conversation* (21.9%) are in between. Among the pronominal references, we observe large differences in the preference for *it* vs. *they* across subcorpora, with numbers ranging from 34.8% *it* in *broadcast conversation* to 87.4% in *newswire*.

Table 13. Reference types per genre in OntoNotes

	it	they	name	noun	other	total
bc	8	15	59	10	13	105
bn	11	12	146	44	12	225
mz	17	11	91	24	4	147
nw	76	11	926	193	36	1242
tc	2	3	7	0	0	12
wb	6	4	52	8	4	74
	120	56	1281	279	69	1805

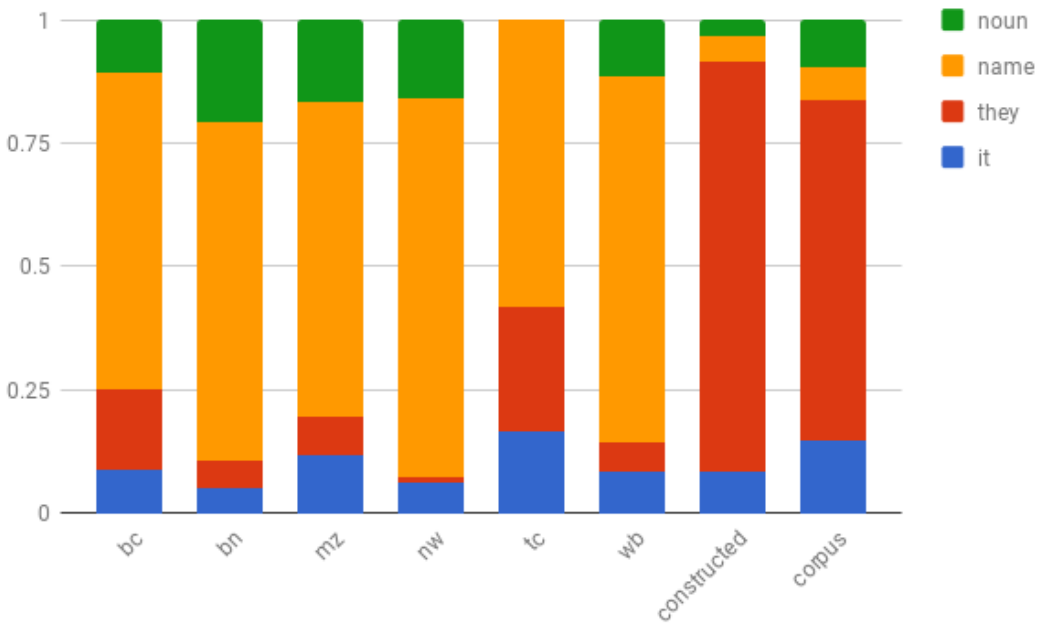

Figure 21. Proportions of reference types. Bars 1–6: OntoNotes (Section 3); Bars 7–8: Continuation studies (Section 5)

6.2.4 The effect of formality

In this section, we examine the hypothesis that the cross-corpus variation in the conceptualisation of organisations as singular or plural can be explained by the different levels of *formality* of the texts.

6.2.4.1 Measuring formality

To measure the formality of discourse, we use an automatic metric proposed by Heylighen and Dewaele (2002). The metric is called *F-score* by the original authors, but we use *Formality score* to avoid confusion with the entirely unrelated F-score derived from precision and recall. The fundamental assumption of Heylighen and Dewaele (2002) is an opposition between *formality* and *contextuality*, with the claim that more formal texts prefer more absolute and less context-dependent forms of expression, which is reflected in lexical choice. The authors identify two (non-exhaustive) subsets of the lexicon that they call *formal* or *non-deictic* and *deictic*, respectively. This distinction is then operationalised via part-of-speech (POS) categories with nouns, adjectives, prepositions and articles taken to

be non-deictic, and pronouns, verbs, adverbs and interjections as deictic. The score is calculated as:

$$F = 100 \cdot \frac{N_{\text{formal}} - N_{\text{deictic}}}{2N} + 50$$

where N_{formal} and N_{deictic} are the counts of formal and deictic tokens and N is the total corpus size.

6.2.4.2 Choice of referring expression

Since the number of pronouns in a corpus enters the computation of the Formality score through the N_{deictic} term, we must exercise care when we measure referential preferences so that we do not use a metric that is correlated by construction with the Formality score. The preference among pronominal references between the conceptualisation of organisation as singular versus collective entities can be measured as the proportion of references with *it* among third-person pronominal references, i.e., N_{it} / (N_{it} + N_{they}). As both *it* and *they* are counted as pronouns in the Formality score, their proportion can be measured independently from the score.

Figure 22 plots the proportion of *it* among pronominal references (x-axis) against the Formality score (y-axis). The ranking predicted by the Formality score seems intuitively reasonable: The *newswire*, *magazine* (two edited written genres) and *broadcast news* (prepared speech in a very formal setting) genres are identified as most formal, whereas *telephone conversations* are predicted to be least formal. The Pearson correlation coefficient between the Formality score and the proportion of *it* vs. *they* is 0.67, which fails to reach significance ($p = 0.146$). However, the Spearman rank correlation reaches a value of $\rho = 0.886$, which is significantly different from 0 ($p < 0.05$). This confirms that formality is a relevant factor to explain the language-internal variation in the number marking of pronouns with organisational antecedents, as was suggested for collective nouns more generally by Hundt (2009).

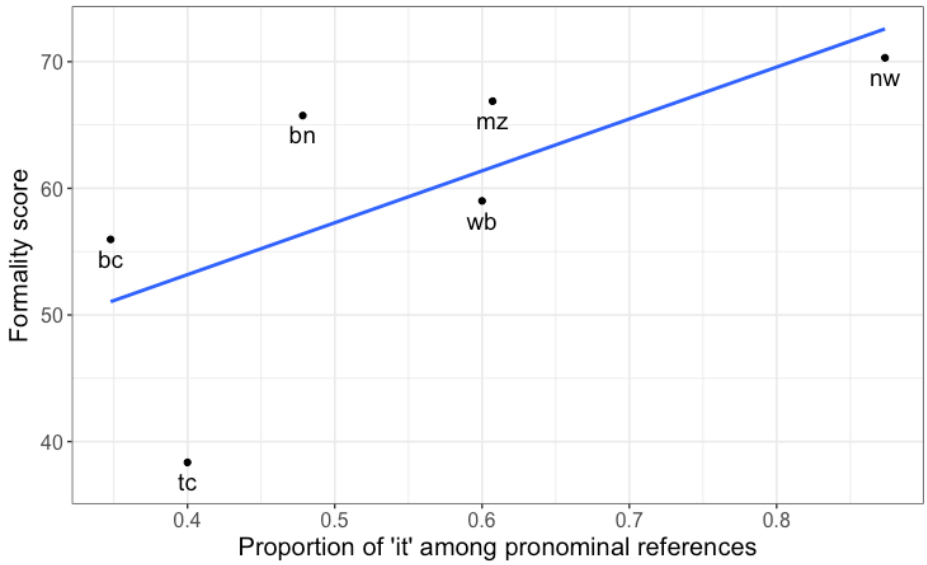

Figure 22. Formality score and prevalence of *it* in OntoNotes subcorpora.

6.2.5 Continuation experiments

Two story-continuation studies presented participants with prompts to elicit entity coreference.

6.2.5.1 Study 1: Constructed stimuli

6.2.5.1.1 Materials

The 16 experimental items consisted of a context sentence and a prompt. The first sentence introduced a named entity in the subject position of the matrix clause with some additional contextual information, followed by a prompt with a discourse adverbial or other connective (e.g., *In the following years*, *Because of this*). The named entities belonged to four categories: names of companies, publishers, sport teams and music bands. The experimental items were interleaved with 48 filler items. They included 20 fillers composed of a sentence introducing two animate or inanimate entities, followed by an adverbial prompt, 24 items for an unrelated production experiment involving the coreference of the pronouns *it* and *this*, and 4 catch trials with a straightforward correct response, which were mentioned in the instructions.

6.2.5.1.2 Participants

Twenty-seven monolingual American English speakers aged 19–63 (mean age 36, σ = 11.2; 15 male) were recruited from Amazon's Mechanical Turk (Munro et al., 2010; Gibson et al., 2011). All had US IP addresses and received $4 for an estimated 30-minute task.

6.2.5.1.3 Procedure

The continuations were collected via a web-based interface that participants accessed directly from Amazon's Mechanical Turk. The website displayed a background questionnaire, a consent form and an instructions page, and then proceeded to display one item at a time with a textbox for participants to write their continuations.

6.2.5.1.4 Analysis

The continuations were annotated for referent type, using the same labels as in the corpus analysis, plus *other* for continuations in which the named entity was not mentioned in any way. One of the authors of the paper annotated the whole set of continuations, and two others labelled half of it. The annotations did not present any real case of disagreement among the authors.

6.2.5.1.5 Results

50 out of 420 continuations were excluded because they were labelled as *other*. This left 370 labelled annotations for the analysis. The results are shown in Table 14 and Figure 21. It is striking that the participants produced an extremely high number of pronominal continuations, most of them with *they* (accounting for 83% of the referential types vs. only 3.1% in the OntoNotes data). By contrast, the *name* category occurred only infrequently (5.1% of types vs. 71% in OntoNotes).

6.2.5.2 Study 2: Corpus stimuli

6.2.5.2.1 Materials

The 24 target passages were extracted from the data described in Section §6.2.3.1. They were interleaved with 76 filler items. 24 of these were extracted from the ParCorFull corpus (Lapshinova-Koltunski et al., 2018). These fillers mentioned a range of referents; the

sentence continuation prompt was an adverbial expression (e.g., *Eventually*). 48 additional fillers were items of the aforementioned unrelated production experiment, and a final 4 fillers repeated the catch trials from Study 1.

6.2.5.2.2 *Participants*

Nineteen monolingual English-speaking participants aged 23–44 (mean age 30, $\sigma = 6.5$; 13 male) were recruited as in Study 1, and received $7 for an estimated 50-minute task.

6.2.5.2.3 *Procedure and analysis*

Identical to Study 1.

6.2.5.2.4 *Results*

43 out of 207 continuations were labelled *other* and excluded from the data set. This left 164 continuations labelled for referential type. The results are shown in Table 14 and Figure 21. While the continuations produced in this study still contain a much larger number of pronominal references than the OntoNotes examples, the proportion of pronouns (83.5%) is considerably lower than in Study 1 (91.6%), and the proportion of *it* among pronominal references is higher (17.5% vs. 9.4%). The difference between the distributions observed in Studies 1 and 2 is statistically significant ($\chi^2 = 145.71$; $p < 0.001$) in a χ^2 test with Monte Carlo simulation (Hope, 1968).

Table 14. Reference types in the continuations

	it	they	name	noun	total
Study 1	32	307	19	12	370
Study 2	24	113	11	16	164
	56	420	30	28	534

6.2.6 Conclusions

Focusing on pronouns referring to proper names, our study confirms a number of results suggested by earlier research concentrating primarily on collective common nouns and verb agreement (Hundt, 2009). There is significant language-internal variation in English in how speakers and writers refer to organisational named entities. In the OntoNotes corpus data, the most frequent way of referring to an organisation is by repeating its name. The number of pronominal references and their distribution among *it* and *they* varies

greatly across genres. As suggested by Hundt (2009), we find a correlation between the level of formality of a text and the prevalence of singular pronominal references.

In the story continuation studies, we observe a distribution of reference types that is more extreme in its preference for *they* than even the most informal OntoNotes genres. This suggests that the patterns we obtain in this type of study are more representative of informal and spoken language than of more formal written genres, despite the written modality of the task. As a result, we cannot automatically generalise the findings from these studies across different genres. However, the combination of crowd-sourced continuation study and corpus analysis provides us with a useful baseline in terms of both methods and results for a planned cross-lingual study of reference to named entities.

6.3 Cross-linguistic preferences in producing reference to named entities[26]

6.3.1 Introduction

Modelling coreference patterns in natural language is extremely complex, as the decisions regarding next mention and referring expressions are influenced by syntactic, semantic, and pragmatic factors. When referring to a named entity, for example, multiple considerations will go into the choice of a referring expression: some are dictated by the grammar of a specific language (e.g., some languages may demand grammatical gender or language marking), others by broader factors implicated in discourse coherence, such as the prominence of the antecedent. Consider the following examples in English (2a), German (2b), and Italian (2c):

2) Example of coreference with *Google* from the experimental data:

a. Google entered the search machine business in 1998. Ten years later, they dominate almost all search on the web.

b. Google trat 1998 in den Suchmaschinenmarkt ein. Zehn Jahre später ist Google an der Spitze der Suchmaschinen.
Google entered the search engine market in 1998. Ten years later, Google is at the forefront of search engines.

[26] This article is currently being prepared for submission by the same four authors of the article in §6.2.

c. Google ha introdotto il suo motore di ricerca nel 1998. Dieci anni dopo, si conferma il primo motore di ricerca a livello mondiale.
Google introduced its search engine in 1998. Ten years later, it remains the top search engine worldwide.

In the example in (2), different choices are made in the different languages, and while speakers of English do not need to commit to a gender, spakers of German and Italian may want to avoid grammatical gender commitment while still appropriately agreeing in number. In English, Google is referred to with the plural pronoun *they*; in German, it is singular (as shown by the 3rd person singular verb *ist*) and it is mentioned by repeating its proper noun, thus not marking gender; in Italian, it is referred to with a null pronoun and the verb agrees in the singular (*si conferma*, "it is confirmed as"), while the gender agrees in the masculine with *motore di ricerca* ("search engine"). Spanish, like Italian, has the possibility to use a null pronoun to avoid gender; French, finally, can only avoid gender by repeating the name of the entity or using the plural pronoun *ils* (a masculine used as a default). These different strategies may have repercussions on the number conceptualisation of named entities; the study here presented attempts to establish baselines in the use of anaphoric reference to named entities across five languages, to inform both linguistic description and computational methods.

Collective nouns, being syntactically singular but semantically (or notionally) plural, are a general problem for coreference applications in computational linguistics (Poesio, 2016). Specifically, the variation of preferences across languages creates an issue for translation, and especially machine translation (MT), as a literal translation of the referring expression found in the input text, even if technically possible, may result in a translation that does not sound natural. Named entities, in fact, often have proper names that do not show clear grammatical number (or gender), and as such they present variable number agreement. It is known that speaker preferences in such contexts can be sensitive to the formality and modality of the context in which the utterance is produced (Hardmeier et al., 2018; §6.2); by adding a cross-linguistic perspective, the following study shows how the choice of expression is also influenced by systemic properties of the language, particularly

whether or not the language offers ways to avoid making a choice about the entity with regard to gender marking, and which these gender avoidance strategies are.

Collective nouns pose to speakers and comprehenders the problem of number agreement, which can be grammatical (singular) or notional/semantic (plural). Named entities, a specific type of collective nouns, add the problem of gender agreement to that of number.

A comparative study is presented investigating the production of referring expressions across five European languages, and how the need for morphosyntactic marking affects speakers' preferences for certain types of referring expressions. In particular, we study how speakers of different languages maintain reference to an organisation (a company, a music band or a sports team) when they refer to it again in the sentence following its first mention in the discourse.

Unlike in English, in languages like French, German, Italian and Spanish nouns are gendered, whether with a binary (masculine and feminine in the three Romance languages) or with a tripartite system (masculine, feminine and neuter in German). This requires a decision by speakers who need to refer to named entities. With common nouns and proper names, a number of factors can have an impact (world knowledge, competence in the language or the noun's membership to a specific inflection class identify its grammatical gender).[27] On the other hand, the proper names of complex entities such as companies, bands or teams are often made up (e.g. *Google*), acronymic (e.g. *IBM*, *IKEA*) or borrowed from names of places or surnames or other not clearly gendered words (e.g. respectively *Amazon*, *Ferrari*, *Juventus*).

Choices regarding grammatical gender have been shown to be dictated for ambiguity avoidance reasons (Hwang, 2020): ambiguity avoidance affects the choice of pronouns, biasing the production towards pronouns that are not gender-ambiguous (and, so, refer-

[27] For example, the two most common morphologic classes of Italian nouns, with singular and plural inflectional suffixes *-a/-e* and *-o/-i*, respectively assign feminine and masculine gender.

entially ambiguous), and even when ambiguity is not present, the choice of referring expressions has been shown to vary with the presence of competitors (Arnold and Griffin, 2007). Nonetheless, choices of grammatical gender or other category to avoid committing to a gendered noun phrase are less studied.

However, the marking of grammatical gender is not mandated in the same way across languages. The marking can be present on the noun itself, in articles, adjectives or on the verb. In Italian for example, articles are mandatory and they mark gender both in the singular and the plural, as do adjectives and past participles of unaccusative verbs; in German, on the other hand, grammatical gender is only relevant to the singular, where it is marked in articles and adjectives, while using a plural would bypass the need for marking. Moreover, a null subject allows the speaker to avoid explicitly marking gender in pro-drop languages like Italian or Spanish: this alone could predict a more consistent use of gender avoidance strategies in these two languages than in German or French. Finally, any of the languages examined except Italian (where proper names of named entities are almost always preceded by a determiner) allows speakers to avoid gender by repeating the proper name of the organisational entity, or to use a hypernymic common noun with a grammatical gender of its own. These features of each language are reported in Table 15.

Table 15. Language features.

	DE	EN	FR	IT	ES
Avoid gender (singular)	−	+	−	+	+
Avoid gender (plural)	+	+	−	+	+
Null pronouns	−	−	−	+	+
Gender priming	−	−	−	+	−
Number priming	+	−	+	+	+

Different strategies of avoiding gender, then, will have different repercussions on the general prominence of the entity: anaphoric referring expressions are used depending on the levels of activation of the antecedent (see e.g. Givón, 1983; Ariel, 1988; 1990; 2004), and repeating a semantically heavier referring expression, which would more commonly be used to introduce a new referent or raise its activation after some time, can incur processing penalties: this dispreference, or even pragmatic inappropriateness, of using a semantically (and phonetically) heavy form to refer to a recently activated antecedent can

be explained by the Repeated Name Penalty (Gordon et al., 1993; 1999), stating that referring to a subject antecedent repeating its noun will cause slower reading times, or in Centering terms (Grosz et al., 1995), where a pronoun is predicted as an anaphor when the centre is maintained, or even with the Gricean model (Grice, 1989), as a violation of the quantity or manner maxims. While null-subject languages like Italian and Spanish could then easily use a null pronoun, following prominence-related biases and also avoiding gender, a speaker of a language like French, where the only gender avoidance strategy is using a heavy referring expression like a noun or a name, may be torn between being spared the commitment to a gender and avoiding any pragmatic oddity brought on by referring to a highly prominent antecedent with an expression more appropriate to less activated referents.

In this article we test how speakers of different languages make referential choices in order to assess (i) whether the avoidance of gender marking biases the number conceptualisation of named entities towards the plural rather than the singular, and (ii) whether the use of semantically (and phonetically) heavy forms such as a repeated proper name or a noun is used as a gender choice avoidance strategy. To this end, a story continuation experiment in English, French, German, Italian and Spanish was used asking speakers to continue sentences about organisational named entities producing the referring expressions of their choice.

The results show a consistently above-chance use of gender-avoiding strategies. Italian and Spanish show a bias towards a singular conceptualisation of named entities (with the notable exception of bands in Italian), while German and English are biased towards plural conceptualisations and French shows a light bias towards the plural. Different types of named entities seem to be conceptualised differently (and consistently across languages). Finally, with regard to the referring expressions chosen in the singular conceptualisation, Italian and Spanish fall on the most minimal side of forms using mostly null pronouns, with German and French using the heaviest set of expressions (mainly names and nouns), thus avoiding grammatical gender despite the processing penalties associated with these choices of form (cf. Gordon et al., 1993), and English using (non-gendered) overt pronouns most often.

6.3.2 Related literature

Perlmutter (1972), noticing that in his dialect collective nouns exhibit singular verb agreement, shows examples that syntactically behave like singulars while semantically denoting a set of elements. To account for the double possibility of syntactic and notional (or semantic) agreement, Corbett (1979) postulates a hierarchy designed with data from a variety of languages, including English, French, German, and Spanish. The hierarchy organises different types of number agreement according to their relative possibility of syntactic agreement: attributes are the most likely to agree syntactically, followed by predicates, relative pronouns, and finally personal pronouns, which are the most likely to agree semantically. Moreover, some factors increase the probability of semantic agreement, in particular the syntactic distance between the controller and controlled elements. The probability of syntactic vs notional agreement is thus determined by a purely syntactic measure.

Evidence has been found for the difference in the number agreement of collective nouns, both in time and among varieties of English (Levin 2001). Diachronically, both British and American English have been showing an increase in singular agreement (Lakaw, 2017). A difference between British and American English is often posited where British English favours plural concord (as in example (2a) above), while in American English collective nouns would display singular agreement (Quirk et al., 1985; Biber et al., 1999). However this assumption has been challenged both with regard to British English (Depraetere, 2003) and to the outer-circle varieties based on British and American English (Hundt, 2006).

Complicating the matter, verbal and pronominal agreements pattern similarly but not identically, with pronominal usage being more likely to accord with the conceptualisation of collective entities as a plurality (Bock et al., 2004; Lakaw, 2017). A possible explanation points to verb agreement being more sensitive to grammatical number and anaphoric agreement being more sensitive to notional number (Corbett 2000), with this dissociation being shown both in language processing and production, even when both types of agreement occur within the same sentence (Kreiner et al.2013). Subject-verb number agreement has also been shown to be affected by how abstract the subject entity is, with more

imageable subjects leading to more plural notional agreements (Eberhard 1999), and by whether the collective noun is modified distributively, which increases its plural construal (Humphreys and Bock, 2005).

Less studied is the question of whether named entities – that is, proper names denoting specific collective entities – pattern differently with regard to morphosyntactic agreement in English: in fact, while collective nouns still have a grammatical number (cf. *company* vs *companies*), proper names do not, and the choice of agreement is left to the entities themselves (where decisions may reflect image and branding reasons: see e.g. Janner, 2018) and to the speakers referring to them. A previous study found a prevalence of the plural conceptualisation of named entities, with 83% choice of *they* as a referring expression but an effect of formality whereby more formal texts showed a higher proportion of singular conceptualisations expressed by the pronoun *it* (Hardmeier et al., 2018). Our study provides a linguistic description of the phenomenon that can keep the formality and genre variables constant, something that would be harder in a corpus study that is not designed specifically to this end.

For Italian, an extensive morphosyntactic study on commercial brand names was conducted by Janner (2018). Gender and number of the entity, mostly indicated by the determiner, were analysed not in relation to an underlying noun which may influence them (e.g. *azienda*, "company", which is a feminine singular), but rather as a systematic, functional variation of meaning. Most brand names show a rigid gender categorisation with only non-significant free variation; however, gender alternation can be used to systematically refer to the company with a feminine singular, to its shares with a feminine plural, and to its product(s) with a masculine form. In French, the feminine plural agreement will be used to talk about the shares of a company (Janner, 2018), and in German too this meaning will be conveyed with a plural, without articles (Grass, 2002). On the other hand, most variation in number conceptualisation is free in Italian, without opposing different meanings or functions.

The proper name of brands is not preceded by an article when a brand is used as an "indefinite plural" (e.g. *il diavolo veste Prada*, “the devil wears Prada”), in enumerations (e.g. *vive*

di pane e coca cola, "she lives on bread and Coca-Cola"), and with some prepositions (e.g. *lavora da/in Google*, "she works at/in Google"), but, most importantly, uses of the determinerless name try to reinforce the commercial value of the name, as often recommended by marketing policies of a company (Janner, 2018: 302ff).

Despite work claiming that notional agreement is not grammatical in French (e.g. Lammert, 2010), plural agreement with collective nouns is indeed possible and it has become diachronically more widespread (Tristram, 2012; Tristram and Ayres-Bennett, 2012), especially with a word clearly denoting a semantically plural entity, such as *majorité* ("majority", Tristram, 2014). However, other collective nouns (such as *foule*, "crowd"), seem to contradict this simple tendency, showing a difference in conceptualisation depending on the article (with determinate nouns agreeing in the singular much more often than indeterminate nouns) and on the presence and number of a complement (Tristram, 2015). This count translate to different predictions when it comes to our study: named entities in French could agree semantically, like collective nouns such as *majorité* and *foule*, or could avoid notional agreement, as predicted by Lammert (2010).

For Spanish, Vande Casteele (2012) notes how the first mentions of named entities are most often accompanied by descriptive elements. It is sensible to assume that these may in turn influence number and gender agreement. Spanish speakers have also been shown to interpret numerosity in a similar way to English speakers despite having a richer morphology (Bock et al., 2012): both languages display effects of distributivity (that is, the distribution of the same attribute over multiple tokens of an object), but pattern similarly both in direction of the biases and magnitude.

In German, effects of distance between antecedent and anaphoric pronouns have been noted: the preferred agreement of pronouns referring to a collective noun is grammatical (singular) more in adjacent sentences than in distant ones, and such agreement is also processed faster across less distance (Schweppe, 2013), as a result of the accessibility of the grammatical number marking decaying over time in working memory. Other studies confirm the validity of the Agreement Hierarchy (Corbett, 2000) in German, with pronouns

being more attracted by the notional number than verbs, in which notional agreement remains non-grammatical or at least rare over time (Birkenes and Sommer, 2015).

In summary, many factors seem to influence the number conceptualisation of collective nouns and named entities, including semantic features of the entity, morphosyntactic factors such as the type and presence of determiner and modifiers, as well as syntactic-pragmatic factors such as distance between the two parts of the agreement and the contextual prominence of the notional number of the entity. The lack of systematic crosslinguistic research makes it unclear whether some of these factors apply crosslinguistically, or if they are specific to the languages in which they were first found. Our study has both a descriptive goal, aiming to find baselines in the use of expressions referring to named entities, as well as exploring gender avoidance strategies that speakers may use in this type of coreference. Moreover, most research focusses on collective nouns, but named entities do not have the same features: most notably, being often made-up words or words in a foreign language, they are not clearly marked for grammatical number, which could lead to notional number being used more frequently for agreement with named entities than for agreement with collective nouns at large.

6.3.3 Experiments

A story continuation experiment was run in each language, using prompts to elicit continuations that would include a reference to a named entity from participants. The design of the study follows the same used in Hardmeier et al. (2018), expanded to cover five languages. The data will be analysed in three ways: analysis of gender avoidance strategies (§6.3.3.3.1), of choice of number conceptualisation (§6.3.3.3.2) and of choice of referring expression (§6.3.3.3.3).

6.3.3.1 Materials

Each language included 18 experimental items, 6 for each type of named entity (company, sport team, music band). Each item was made of a context sentence introducing a named entity as its subject as well as contextual information, followed by a continuation prompt (either adverbial, e.g. *Recently*, or other connective, e.g. *The following year*). All verbs used past tense: past simple in English (making it possible to not prime the entity's number,

while in the present or present perfect it would have been indicated in the conjugation of the verb), *präteritum* in German and preterite in Spanish (both equivalent to a past simple), *passé composé* in French and *passato prossimo* in Italian (both equivalent to a present perfect, the most natural choice in these two languages).

An example item in the five languages is given in (3). Some of the items, for named entities renowned on a global scale like the one in (3), were equivalent in each language, while others used named entities and context likely to be known by each language's participants, to ensure natural continuations informed by world knowledge.

3) Example experimental item in the five languages:
    a. **EN**: Google entered the search machine business in 1998. Ten years later,
    b. **FR**: Google est entré dans le secteur des moteurs de recherche en 1998. Dix ans plus tard,
    c. **DE**: Google trat 1998 in den Suchmaschinenmarkt ein. Zehn Jahre später
    d. **IT**: Google ha introdotto il suo motore di ricerca nel 1998. Dieci anni dopo,
    e. **ES**: Google entró al mercado de los motores de búsqueda en 1998. Diez años después,

The items were interleaved with a total of 48 non-target items: 24 items for an unrelated study about reference to events versus (simple) entities, composed of a sentence with one or two entities followed by a sentence to continue after a pronominal prompt, and 24 fillers composed of a sentence introducing two referents and an adverbial prompt starting the sentence to be continued. 4 of the fillers were constructed as catch trials with more obvious responses, and were mentioned in the study's instructions to give participants an incentive to focus on the task.

While some of these languages prime the participants for gender (e.g. Italian) or for number (all languages but English) through the verb inflections and/or articles in the stimuli, priming did not seem to be a relevant effect, with most of the continuations in fact going in the opposite direction than the priming. The effect of priming is thus not considered in the analysis in §6.3.3.3.

6.3.3.2 Participants and procedure

Participants were recruited on Amazon Mechanical Turk targeting users from specific IP addresses (i.e. USA, France, Germany, Italy and Spain). When MTurk did not provide enough participants, the remaining participants were recruited among students from the same countries at the University of Edinburgh. The task was estimated to last 45-60 minutes and all participants were compensated with $8/€7/£7. Summary statistics of the participants are reported in Table 16.

Continuations were collected via a web-based interface that participants could access from their own computer through MTurk. On-campus participants were sent a link to access the experimental interface. The website displayed a background questionnaire, a consent form and a page with instructions. Then, each item was presented on a page by itself with a text box for participants to write their continuations.

Table 16. Summary statistics regarding the participants.

	N	Age: range	mean	σ
English	46	22–70	37.1	11.3
French	53	18–55	30.9	10.0
German	46	18–55	32.0	10.9
Italian	53	18–48	29.7	8.3
Spanish	51	18–67	33.0	9.7
	248	18–70	32.4	10.2

6.3.3.3 Annotation, analysis and results

The continuations were annotated by the authors for referring expression and number agreement on the verb. Specifically, the continuations were coded for the repetition of the proper noun, the use of a common name (e.g. *the company*), or the use of a pronoun. Pronouns were further divided into null and overt pronouns in Italian and Spanish, and impersonal continuations were excluded because they were only found in German and French in very low numbers (respectively 14 occurrences of *man* and 1 of *on*). Continuations not including a named entity were also excluded.

6.3.3.3.1 *Analysis of gender avoidance*

The first analysis concerns the rate at which participants chose a gender-avoiding form to refer to a named entity. English is excluded from the analysis as it does not have grammatical gender. In all the four considered languages, both the repetition of the named entity's name and the use of a common noun were considered gender avoidance strategies. Moreover, in Italian and Spanish null pronouns were coded as gender avoiding, as were the plural pronoun *sie* in German and the plural pronoun *ils* in French.[28] The percentage of times participants chose to use a gender avoidance referring expression is visualised in Figure 23.

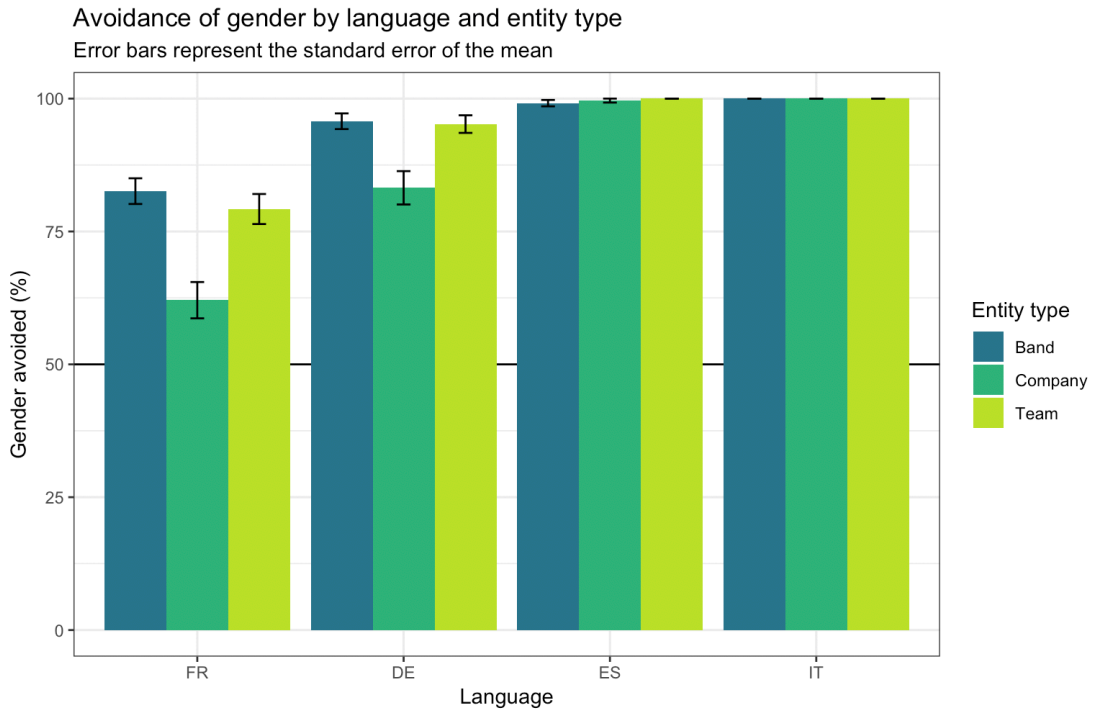

Figure 23. Gender avoidance in French, German, Spanish and Italian.

[28] Technically, the pronoun *ils* is the plural masculine pronoun (with the feminine being *elles*). However, since a group containing at least one masculine entity is referred to with *ils* as a general plural pronoun, *elles* would not be a possible choice in the context of this study – and, in fact, it never appeared in the participants' continuations.

Gender avoidance was modelled using generalised mixed-effects logistic regression (Bates et al., 2015). The model predicted the GenderAvoided binary outcome from the two predictors of language and entity type, with random intercepts by participant and by item. English was excluded from the model because it does not have grammatical gender, and Italian was excluded as grammatical gender was always avoided, either with a null pronoun (96.3%) or nouns and names (respectively 2.2% and 1.4%). The predictors were centred as differences, as "DE–FR" and "ES–DE", and "Company–Band" and "Team–Company". The results of the model are reported in Table 17.

Table 17. Use of gender avoidance strategies: Estimated model fixed effects.

Effect	Estimate	Std.Error	z value	$Pr(>\|z\|)$	
(Intercept)	3.56	0.26	13.79	< 0.001	***
DE−FR	1.51	0.33	4.57	< 0.001	***
ES−DE	3.38	0.67	5.09	< 0.001	***
Company−Band	−1.42	0.23	−6.20	< 0.001	***
Team−Company	1.20	0.23	5.27	< 0.001	***

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

The results of the model show a significant difference between languages, wherein Spanish participants avoid gender more than German ($p < 0.001$) and German participants avoid gender more than French ($p < 0.001$). The fact that Italian and Spanish would avoid gender the most can be predicted by the fact that their null pronouns are both gender avoiding and the preferred forms for topic continuation. Moreover, gender is avoided more in bands and teams than in companies (both $p < 0.001$).

6.3.3.3.2 Analysis of number conceptualisation

A breakdown of how named entities are conceptualised is reported in Table 18 and a visualisation is provided in Figure 24. As can be noticed, in French and German nouns and names sway the number conceptualisation towards the singular: in the other languages, the effect is less pronounced, as these referring expressions are less used (see Figure 25).

Table 18. Number conceptualisation in the five languages (%).

	DE	EN	FR	IT	ES
	Names, nouns and pronouns				
Sing	37.7	17.3	56.9	55.9	78.0
Plur	62.3	82.7	43.1	44.1	22.0
	Pronouns only				
Sing	11.2	13.4	36.6	54.5	76.4
Plur	88.8	86.6	63.4	45.5	23.6

The analysis of singular vs plural conceptualisation only included pronouns (of any type). The choice of number was modelled using generalised mixed-effects logistic regression (Bates et al., 2015), predicting the isSingular binary outcome using language and entity type as predictors as well as their interactions (which improved the model fit, $p < 0.001$). All predictors were centred as differences (as in §6.3.3.3.1). Furthermore, random intercepts were included for participant grouped by language and for item grouped by entity type. The results of the model are reported in Table 19.

Table 19. Number conceptualisation: Estimated model fixed effects.

Effect	Estimate	Std.Error	z value	$Pr(>\|z\|)$	
(Intercept)	−0.54	0.12	−4.31	< 0.001	***
Company−Band	2.98	0.26	11.27	< 0.001	***
Team−Company	−0.92	0.17	−5.48	< 0.001	***
EN−DE	0.46	0.38	1.20	0.23	
EN−DE : Company−Band	−0.67	0.61	−1.10	0.27	
EN−DE : Team−Company	0.19	0.60	0.32	0.75	
FR−EN	2.56	0.32	7.87	< 0.001	***
FR−EN : Company−Band	−0.42	0.45	−0.94	0.35	
FR−EN : Team−Company	0.82	0.46	1.79	0.07	
IT−FR	−0.51	0.45	−1.14	0.25	
IT−FR : Company−Band	7.99	1.11	7.19	< 0.001	***
IT−FR : Team−Company	0.60	0.50	1.19	0.24	
ES−IT	1.67	0.45	3.71	< 0.001	***
ES−IT : Company−Band	−8.08	1.11	−7.31	< 0.001	***
ES−IT : Team−Company	0.43	0.49	0.88	0.38	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

The model shows a significant effect of the type of entity where companies are conceptualised in the singular more than both bands and teams (both $p < 0.001$). While no significant difference is seen between German and English, the latter is significantly different from French, which uses more plurals ($p < 0.001$). The difference between French and Italian is only significant in the interaction with "Company–Band" ($p < 0.001$), meaning that the difference between the two languages is most relevant in how they treat these two types of named entities (as can be seen in Figure 24), where the difference between bands and companies in Italian is much wider and bands are always conceptualised with a plural. The difference between Italian and Spanish, finally, is significant both as a main effect where Spanish conceptualises in the singular more ($p < 0.001$), and as an interaction with "Company–Band" ($p < 0.001$), with the difference between these two entities being larger in Italian, not unlike with the French comparison.

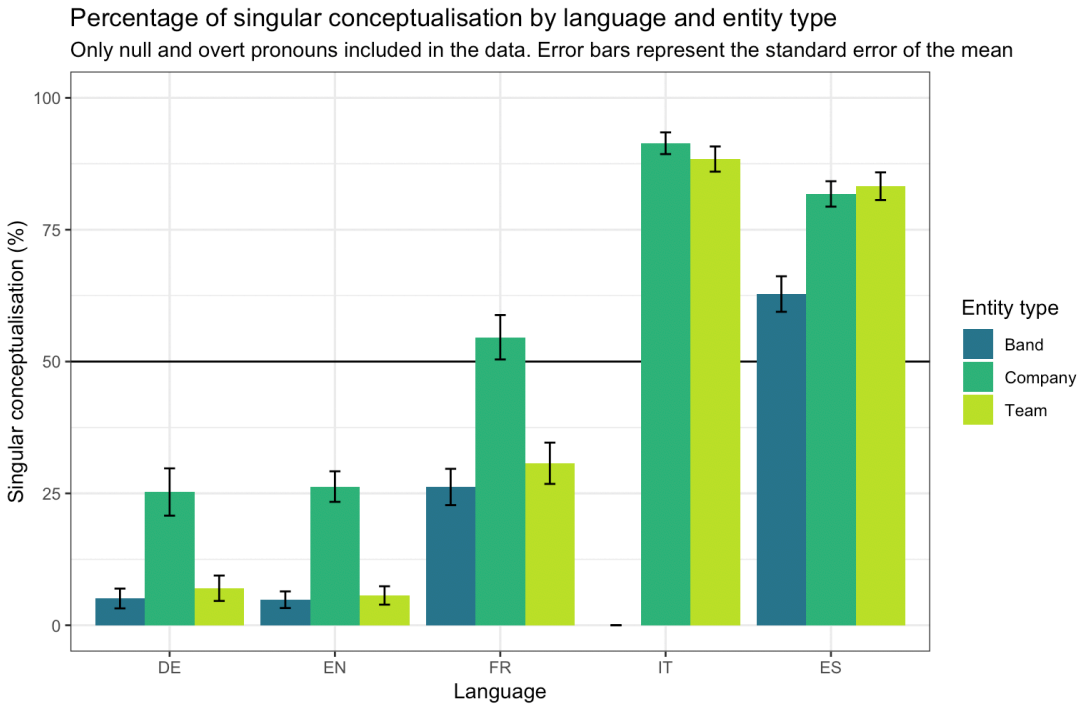

Figure 24. Incidence of singular conceptualisation by entity type and language.

6.3.3.3.3 Analysis of referring expressions in the singular conceptualisation

The analysis of referring expressions focusses on the singular conceptualisation only: in fact, when choosing the plural, participants overwhelmingly decide to use pronouns

(overt or null depending on the language), doing so 99.6% of times; moreover, common nouns would not be used in the plural as they would refer to multiple named entities rather than just the one. When referring to an entity in the singular number, conversely, a more varied set of forms is used, and the referring expressions participants chose can be put on a scale from semantically lightest to heaviest: null pronouns, overt pronouns, common nouns and finally the entity's proper name. Since the continuations refer to the topical subject entities of the preceding prompts, lighter referring expressions would normally be expected (e.g. Ariel, 1990). However, avoidance of gender markings could lead speakers to disregard the usual expressions and the bias placed on them by prominence. These choices are visualised in Figure 25.

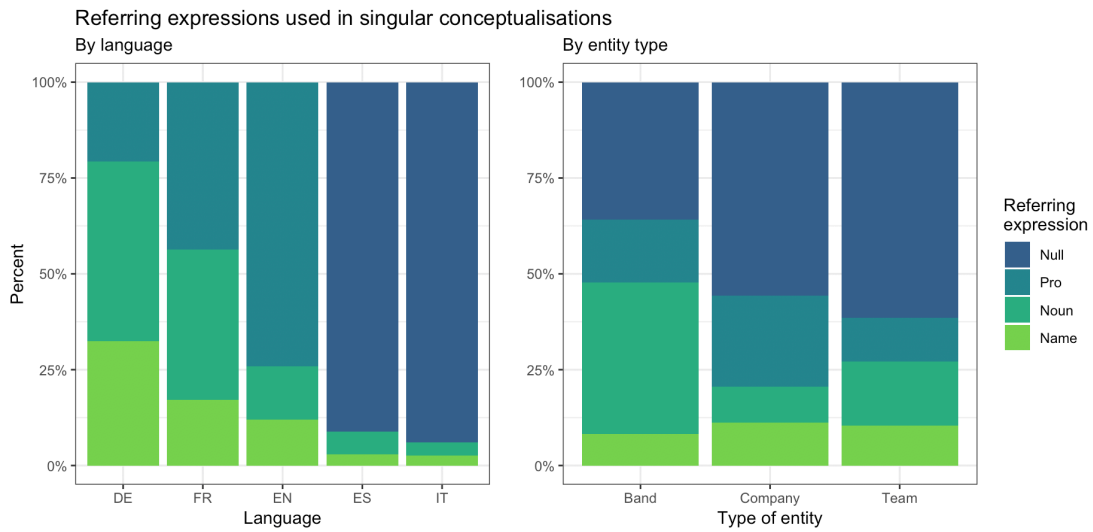

Figure 25. Referring expressions used when a singular conceptualisation is chosen, by language and entity type.

The choice of referring expression in the five languages was modelled with a generalised additive model for ordinal responses (Wood, 2011), placing the referring expressions in a gradient of semantic intensity (1 = null pronoun; 2 = overt pronoun; 3 = noun; 4 = name). The model predicted the Referring expression from the language and type of entities, with smooth terms (random effects) for participant grouped by language and item grouped by entity type. All factors were centred. The results of the model are reported in Table 20. In the table, each line is to be read as the difference between two categories: for

example, the difference between French and German gives a negative estimate, showing that German has heavier referring expressions than French, and so on.

The results of the model show a significant difference between French and German ($p < 0.001$), between English and French ($p = 0.001$), and between Spanish and English ($p < 0.001$), confirming the fact that German uses heavier referring expressions, followed by French, English and finally null-pronoun languages Spanish and Italian, as visualised in Figure 25. The difference between Italian and Spanish is not significant ($p = 0.17$). Moreover, the results show that Bands are referred to by generally heavier referring expressions than Companies ($p = 0.002$), but note that this effect is likely driven by the fact that in Italian bands are always plural (see Figure 24), and thus Italian is not contributing any null pronouns to this singular subset of the data. Teams also elicit heavier forms than Companies ($p = 0.04$). This last effect is probably due to the higher incidence of nouns in the Teams type than in Companies (also visualised in Figure 5).

Table 20. Choice of referring expression in the singular conceptualisation: Estimated model parametric coefficients.

Effect	Estimate	Std.Error	z value	$Pr(>\|z\|)$	
(Intercept)	−0.47	0.07	−6.50	< 0.001	***
FR−DE	−0.81	0.16	−5.05	< 0.001	***
EN−FR	−0.63	0.19	−3.26	0.001	**
ES−EN	−4.00	0.23	−17.63	< 0.001	***
IT−ES	−0.38	0.27	−1.39	0.17	
Company−Band	−0.47	0.15	−3.10	0.002	**
Team−Company	0.31	0.15	−2.04	0.04	*

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

6.3.4 Discussion

The results of the study show both similarities and differences across languages: all languages use gender avoiding strategies when referring to named entities (except English because gender does not apply to it; Figure 23), which leads the speakers to choose specific referring expressions: this in turn causes some languages to be unbalanced towards lighter forms, like Italian and Spanish null pronouns, or heavier referring expressions, most notably in German where the repetition of names and nouns is consistent. The ratios

to which languages choose singular or plural conceptualisation of named entities also varies (Figure 24), with German and English showing the most plural conceptualisation, that is closer to a notional interpretation of number, and Spanish and Italian showing more singular conceptualisations. While different types of entities pattern similarly across languages, with companies being interpreted as a singular entity more often than bands and teams, some grammatical idiosyncrasies can be seen, specifically in the Italian treatment of bands, where a singular agreement is ungrammatical.

The fact that Italian and Spanish mostly use null pronouns, both in the singular and the plural, is unsurprising: regardless of the number conceptualisation, this is the most economical referring expression, and its use to refer to a previously topical entity is in line with the predictions of Accessibility Theory (Ariel, 1990) and Centering Theory (Grosz et al., 1995). It is, then, a very convenient way to refer to a named entity: not only is it the most pragmatically appropriate expression, but it also avoids making a choice with regard to the gender of the entity.

The lesser use of gender avoiding strategies in French and German follows from the same principles. The repetition of a name would incur the Repeated Name Penalty (Gordon et al., 1993), and even the use of a noun goes against models that predict "accessibility ~ reduction" (e.g Ariel, 1990; Givón, 1983). The choice of a plural conceptualisation of named entities in German prevents these pragmatic infelicities by using the non-gender marked pronoun *sie*, which is used in over 99% of cases (the rest are two repeated names). The mostly notional number agreement in German is somewhat surprising when considered alongside the evidence about collective nouns, in which the prevalent form of number agreement is grammatical (e.g. Schweppe, 2013; Birkenes and Sommer, 2015). However, named entities do not have the morphological determinacy that collective nouns have: while collective nouns have well-defined grammatical number and gender, named entities do not, and this could explain the higher prominence their notional number has when reference is made to them in a subsequent sentence.

The same applies to French: names and nouns are used in over half of cases in the singular, and in the plural the pronoun *ils* is the only choice. The French results confirm that, contrary to the claim in Lammert (2010), notional agreement for anaphoric pronouns is indeed used. This replicates the agreement patterns seen for collective nouns in French (e.g. Tristram, 2012). As explained by the Agreement Hierarchy (Corbett 1979), this does not automatically translate to the possibility of notional agreement with the verb in French, which should be investigated separately.

As for English, the most used referring expression in the present is the neuter pronoun *it* (74%, with the rest being equally distributed between names and nouns); in the plural, the overwhelming form is *they*. The heavy bias towards a plural conceptualisation, moreover, replicates our previous results (Hardmeier et al., 2018). Interestingly, English leans towards a notional, plural agreement, with an almost identical pattern to German when looking at the different types of named entities. English is obviously different from German in that it does not have grammatical gender (which can be seen in the choice of referring expressions in the singular, Figure 25). Again, the notional agreement of anaphoric pronouns can be explained with reference to the Agreement Hierarchy (Corbett 1979). Moreover, collective nouns in English have been showing a diachronic increase in singular agreement (Lakaw, 2017), and classic grammars of English report that the American dialect (the one we report data on, targeting participants from the USA) also tends to more singular agreement than other varieties of English (Quirk et al. 1985; Biber et al., 1999).

Finally, the fact that companies are generally more conceptualised using notional number ("Ikea" as plural) than bands and teams could relate to the fact that while music bands and sports teams have a smaller number of often individually known members, companies are made of a very large number of usually anonymous members. This patterns with findings on abstraction in English, showing that the notional number of an antecedent is more prominent and accessible in less abstract and more imageable subjects, leading to more plural agreements (Eberhard, 1999), and extends this observation to other languages.

6.3.5 Conclusions

The study targets named entities as a particular case of collective nouns. Alongside the conflict between grammatical and notional number, speakers have to add the indeterminateness of these names' grammatical gender. Targeting three different categories of named entities in five different languages, the results show that, crosslinguistically, the type of entity biases its number conceptualisation, and in particular that companies (being composed of more numerous and abstract members) tend to be conceptualised in the singular more often than music bands and sports teams.

Another finding regards two conflicting linguistic behaviours: the avoidance of gender-marked referring expressions to refer to named entities, and at the same time the attempt to be pragmatically felicitous by using as light a referring expression as possible. Italian, Spanish, French and German speakers all consistently avoided gender, in different ways: Italian and Spanish using null pronouns, and French and German using a plural conceptualisation with pronouns or a singular conceptualisation with nouns and names. A lower frequency of use of gender avoiding referring expressions could be seen in French and German when compared to Italian and Spanish, in line with the fact that French and German's gender-avoiding referring expressions are also not perfectly felicitous from a discourse pragmatics point of view (see e.g. the Repeated Name Penalty, Gordon et al., 1993).

The number conceptualisation of named entities was shown to mostly pattern similarly to that of collective nouns, with Spanish and Italian on the most grammatical (singular) side of the spectrum and German and English on the most notional (plural). Further research will be needed to tease apart in more detail the various factors leading to the choice of number conceptualisation and of a specific referring expression, relating them to the prominence of different morphosyntactic, semantic and pragmatic features of their antecedents.

6.4 Conclusions

In this chapter, two studies about pronominal coreference with named entities in five languages were reported. Story continuation experiments were used to examine how partic-

ipants would refer to named entities of different types: the choice of number conceptualisation and the referring expressions participants chose were analysed. Furthermore, a corpus study was run on English to check for variations in the conceptualisation across genres of different formality. The number conceptualisation of named entities in anaphoric patterns point to the relative prominence of two conflicting aspects of named entities: grammatical number, that is a singular representing the entity as a single object, and notional number, a plural understanding the entity as the sum of its parts.

In the first study (§6.2), English was seen to corefer to named entities mostly notionally, in the plural, with the pronoun *they*. However, an effect of formality could also be seen whereby more formal genres show a larger occurrence of singular agreement with the pronoun *it*. The aim of the second production study (§6.3), then, was to show which pattern different languages take, and whether this is related to systemic properties of the languages such as their morphosyntactic marking of number and gender, and whether smaller entities, which are more clearly imageable and defined by the sum of their parts, bias number conceptualisation towards the plural (as shown outside the context of named entities by Eberhard, 1999). Mandatory gender marking poses a choice to participants, who can either decide to assign a gender to the entity or to use a referring expression (and thus conceptualisation) that avoids the choice. However, choosing to avoid gender may lead to breaking other discourse principles, such as not referring to a prominent entity with a heavy referring expression (e.g. Ariel, 1990).

The results of the study show very different patterns crosslinguistically, with Italian and Spanish being the most biased towards singular conceptualisation: in fact, both languages have the possibility to avoid gender with a null pronoun. In this case, the lightest form coincides with the one avoiding the choice of gender, and the grammatical number of the named entities is respected. In German and French, on the other hand, the name of the entity (or a noun denoting its category, e.g. *company*) is repeated, despite the Repeated Name Penalty (Gordon et al., 1993). Finally, German and English occupy the most notionally coreferring end of the scale. The effect of the type of named entities confirmed the imageability effect seen by Eberhard (1999), with music bands and sport teams coreferring more often than companies using a plural conceptualisation.

These results are evidence that the prominence of the grammatical vs notional number of a named entity is both language specific, as shown in the different distribution of the languages, and cross-linguistic, as seen in the distribution of the different types of entities and the effect of their imageability. Moreover, the choice seems to change across levels of formality, with a likely influence of prescriptivism or editorial guidelines. This, added to variation of the number choice over time and space (e.g. Levin, 2001) points to an extremely complex interplay of factors biasing whether named entities are preferably seen as a holistic unity or the sum of their parts.

Chapter 7 Conclusions

7.1 Context of the research

In this thesis I have examined phenomena related to prominence as defined and displayed in different areas of language. This exploration followed a path zeroing in on subsequently more minute components, starting from the outward aspect of prominence in language's superficial form, to the relation between prominence and coreferential patterns, to the prominence of an event versus the entities involved in it and, finally, to the prominence of different conceptions of an entity. At the different levels, a crosslinguistic approach was taken, comparing the effects of prominence in English, French, German, Italian, and Spanish.

Prominence is defined both linguistically and cognitively in multiple theoretical and empirical frameworks, and the concept often overlaps with other ideas such as those of accessibility, salience, activation, or emphasis. The notion of discourse prominence has been formalised starting from the terms above (see e.g. Von Heusinger and Schumacher, 2019); however, its resistance to definitions and manifold manifestations in the literature do not need to be a hindrance to looking at prominence at different levels of linguistic analysis. As a phenomenon, it is intuitively experienced at least at the discourse level: most speakers of a language would notice which of its constructions are marked and, although maybe not explicitly, they would be able to recognise their function.

This thesis then considers the relationship between prominence and coreference at various levels, comparing the results of an empirical method with predictions from general theories. It takes a crosslinguistic perspective and one that does not assume a parallel between language interpretation (which has been the focus of much psycholinguistic work, in particular relating to processing) and language production. With the interstitial existence of prominence as both a cognitive and linguistic force, the question is how it affects speakers in different contexts and in different languages.

Many theoretical frameworks can be used to make predictions about how prominence may influence coreferential patterns and/or the choice of anaphoric referring expression, although not always to a precise extent. Starting from the pragmatic and syntactic analyses of Grice (1989) and Chomsky (1981), theories have explained the relationship between coreference and prominence purely pragmatically (the Neo-Gricean accounts; §2.2) or purely syntactically (e.g. Carminati, 2002; §2.5), focussing on accessibility and how it allows/demands reduced referring expressions for anaphoric relations (§2.3), portraying prominence as a cognitive function related to short term memory (§2.4), or, finally, in Bayesian probabilistic terms (Kehler et al., 2008; §2.5).

The first half of this thesis looked at prominence in its external, "formal" aspect in English (Chapter 3), and at how it influences coreference choices in Italian (Chapter 4). Both studies used the construction of the cleft sentence as a case study of how prominence can be signalled syntactically and pragmatically. This construction made it possible to check whether different markings have been (re)assigned to new functions or types of emphasis, and whether predictions from well-known theories mostly built on English hold up when applied to a pro-drop language like Italian.

The second half moves from prominence in discourse to prominence in meaning, considering the coreference patterns of entities versus the events in which they are involved and how the prominence of these two types of referents influences both the choice of antecedent and referring expressions (Chapter 5) and, at a lower level, whether grammatical or notional number is more influential when choosing the number conceptualisation of a named entity (Chapter 6).

7.2 Main findings

In the first study (Chapter 3), the aim was to explore how prominence is expressed in English, whether different types of prominence markers exclude each other or stack up in the same sentence, and whether there is systematic variation in the markings for stylistic reasons. Cleft sentences were chosen as a case study because they are a very marked way to express prominence and contrast syntactically, and because they have parts whose choice is optional, namely contraction of the copula (*it's* vs *it is*), case of the pronouns (*she* vs *her*)

and complementiser (*that* vs *who*). Sentences including these different features were mined from Twitter, and the results showed systematic associations between these features, with the uncontracted copula, nominative pronoun and *who* being used concomitantly to mark discourse prominence or emphasis. This is in keeping with predictions from theories such as Accessibility Theory (Ariel, 1990; see also §2.3 in general), whereby more marked forms are used to refer to more marked referents: in fact, the uncontracted copula is phonetically heavier than the contraction, the nominative form of the pronoun is rarer and perceived as formal, and *who* is semantically richer than *that*. A further comparison between the Twitter dataset and data from iWeb, a corpus of general-use web language, shows significant differences in levels of emphasis based on their formality, with Twitter displaying a wider range in its signalling of emphasis.

After looking at the external aspect of prominence, Chapter 4 presents a study in Italian concerning how prominence (again, marked by cleft constructions) influences the choice of antecedent and referring expression in language production. A story continuation study presented participants with clefts of the subject or the object, and asked participants to continue the text either within or between sentences. The results point to a complex interplay between syntactic role, information structure and coherence relations whereby antecedents marked as contrastively prominent are, somewhat counterintuitively, less often picked up for next-mention but, when chosen as antecedents, they are referred to with lighter referring expressions (such as the null pronoun). Choice of referring expression was also shown to change with sentence boundary, with between-sentence anaphora prompting heavier referring expressions. These results have consequences for two main assumptions often found in the literature on coreference, namely that the interpretation and production of referring expressions mirror each other, and that in a null-subject language like Italian null and overt pronouns will exhibit a division of labour, with an association of the null pronoun with the most salient referent. These assumptions conflate two distinct properties of coreference: the hearer's estimate of *who* the speaker will mention, and that of *how* the referent will be mentioned; and prominence was shown to affect these estimates in dissimilar ways. The new results show that marking

a referent as a focus will be detrimental to its next-mention probability, but will bias the antecedent's anaphoric expression towards lighter forms.

In the second half of the thesis, prominence is explored with respect to its effects on smaller components of meaning. Chapter 5 considers event structure, and how reference is made to either an event or to the entities involved in it, exploring the question in five languages (English, French, German, Italian, and Spanish). Two interpretation studies are presented in which the accessibility and complexity of an entity is manipulated by changing features of the verb, such as its status in the causative-inchoative alternation and its aspect. The referring expressions were also manipulated. The results showed consistent cross-linguistic biases: personal pronouns are biased to corefer with entities, while demonstratives prompt more event coreference. Because events are more complex than entities, this result can be explained cognitively (cf. theories in §2.4) or by theories predicting reduction from the accessibility of the antecedent (§2.3). However, this does not manifest as a clear division of labour: instead, the results showed a more gradable difference where more factors come into play. In fact antecedent complexity increased the rate of event coreference, likely by diminishing the prominence of the explicit entity by making another, implicit, entity available.

Finally, named entities were used in Chapter 6 to explore the prominence of different components of meaning: that is, whether the unity or plurality aspect of a named entity is more prominent and, thus, whether its grammatical or notional number determines the agreement of a next anaphoric mention. This choice was seen to vary greatly crosslinguistically, with Italian and Spanish forming agreement grammatically most often, and German and English being more biased towards notional agreement. This choice was related to the use of strategies to avoid grammatical gender marking, which is mandatory in all of the tested languages except English and, at the same time, is not clearly marked on the collective entities' proper names. Furthermore, the type of entities also made a difference in their conceptualisation, with entities in which the individual members are more prominent and imageable (like music bands and sports teams as opposed to companies) being more likely to agree in the plural. Finally, the agreement of the pronoun was shown to be

influenced in English by the level of formality of a text, with informal speech displaying more notional agreement.

7.3 Implications of the research and future directions

The four lines of research presented in this thesis have made it possible to show some patterns arising in coreference because of prominence. In particular, three main findings can be outlined.

First, the investigation of referring expressions and formal variants in general showed how divisions of labour are most often not absolute, if they are present at all: from the genre- and formality-influenced distributions found in Chapters 3 and 6, to the functional divisions between lighter and heavier forms in coreference seen in Chapters 4 and 5, the choice of referring expressions seems to be not only influenced by the need to refer to a specific antecedent, but additionally by factors ranging from the stylistic to the morpho-syntactic. Considered as a way of grouping these factors, prominence only begins to tap into the complexity of relations (both linguistic and extralinguistic or cognitive) that go into the choices of antecedent and referring expressions.

Second, throughout the research and particularly in Chapter 4 it is evident how the interpretation and production of anaphoric expressions cannot be reduced to a Gricean parallel of causes and effects, where following maxims or rules carries the same effects on both sides of the communication. This is somewhat related to the absence of a clean-cut division of labour, and goes to show how speakers do not simply plan their utterances collaboratively to be understood, but also make choices dictated by the convenience and ease of producing language.

Third, with prominence being often explained as not only a linguistic but also a cognitive phenomenon, a recurring question is whether language-specific theories can be easily expanded to cover multiple languages. Both similarities and differences can be seen across languages: for example, Chapter 5 showed a general crosslinguistic tendency for heavy referring expressions to refer to more complex antecedents, but Chapters 4, 5 and 6 showed the unfeasibility of a straightforward application of English-based theories to Italian coreference patterns and the crosslinguistic variation in conceptualisation and form. While

some hypotheses are given to explain this variation, more work is needed to tease apart which factors are language-specific and which are more universally related to language processing and production.

Finally, future research on prominence and coreference has the potential to be extended to multiple applications, other than linguistic theory. A more precise understanding of the natural patterns displayed by language as a whole or by a specific language can be used in the assessment of computational systems and in the evaluation of their output, with a case in point being machine translation: since many of the choices of antecedent and referring expressions display some optionality, evaluation can and should take into consideration not only the grammaticality but also the appropriateness and naturalness of an output. On the psycholinguistic and psychological side of research, an interesting step to assess the cognitive relevance of prominence would be to measure its effects in online processing and productions, applying methods such as the measurement of reaction times, eye tracking, or neurolinguistic measures, with the ultimate aim of finding whether more fine-grained patterns can be seen in the mind than can be accounted for by speakers' patterns of choices.

Appendix A: Twitter search queries (cf. §3.2)

This appendix lists the queries used to mine Twitter for cleft sentences through TAGS (Hawksey, 2016):

- Lines 1–8 search for accusative pronouns, 9–16 for nominative pronouns;
- Lines 1–4 and 9–12 search for *that*, 5–8 and 13–16 for *who*;
- Lines 1, 2, 5, 6, 9, 10, 13 and 14 search for uncontracted clefts, the others for contracted clefts;
- Odd lines search for feminine pronouns, even lines for masculine pronouns.

```
1.  %22it%20is%20her%20that%22 AND -filter:retweets
2.  %22it%20is%20him%20that%22 AND -filter:retweets
3.  %22it%27s%20her%20that%22 AND -filter:retweets
4.  %22it%27s%20him%20that%22 AND -filter:retweets
5.  %22it%20is%20her%20who%22 AND -filter:retweets
6.  %22it%20is%20him%20who%22 AND -filter:retweets
7.  %22it%27s%20her%20who%22 AND -filter:retweets
8.  %22it%27s%20him%20who%22 AND -filter:retweets
9.  %22it%20is%20she%20that%22 AND -filter:retweets
10. %22it%20is%20he%20that%22 AND -filter:retweets
11. %22it%27s%20she%20that%22 AND -filter:retweets
12. %22it%27s%20he%20that%22 AND -filter:retweets
13. %22it%20is%20she%20who%22 AND -filter:retweets
14. %22it%20is%20he%20who%22 AND -filter:retweets
15. %22it%27s%20she%20who%22 AND -filter:retweets
16. %22it%27s%20he%20who%22 AND -filter:retweets
```

Appendix B: Raw data numbers (cf. §3.2)

Table 21. Raw numbers of the feature variants in the two Twitter data collections.

First round of data collection	
CASE	
Nominative	311
Accusative	487
CONTRACTION	
Contracted	485
Uncontracted	313
COMPLEMENTISER	
That	286
Who	512
REFERENTS NO.	
One	650
Two	148
Clefted subject	112
Clefted object	36
EMPHASIS MARKERS	
No Emphasis	685
Emphasis	113
Exclamation	55
Capital pronoun	33
Capital pro. + !	3
Capital initial	4
Capital cleft	2
Asterisks	4
Nonstandard comma	6
Nonst. comma + !	1
Other	5
Total number of clefts	798

Second round of data collection	
CASE	
Nominative	242
Accusative	374
CONTRACTION	
Contracted	405
Uncontracted	211
COMPLEMENTISER	
That	256
Who	360
Total number of clefts	616

Appendix C. Results with liberal annotation (cf. §5.3)

In this appendix, we report the results of the models run with the liberal annotation (Table 10). The models were chosen following the same procedure described in §5.3.5. Other than the fixed effects as shown on the table, the models all included a random intercept for condition by participant and a random intercept by item, except for the French model which only included the two intercepts and the Italian model which only included a random intercept by participant.

Table 22. Estimated models fixed effects (liberal analysis).

Effect	Estimate	Std.Error	z value	$Pr(>\|z\|)$	
		English			
(Intercept)	0.38	0.14	2.78	0.005	**
Condition (this)	3.70	0.34	10.76	< 0.001	***
Aspect (perf)	-0.28	0.19	-1.46	0.14	
Alternation (alt)	0.74	0.20	3.73	< 0.001	***
Aspect:Alternation	-1.85	0.42	-4.40	< 0.001	***
		French			
(Intercept)	0.27	0.12	2.30	0.02	*
C'est − Cela	-0.04	0.21	-0.20	0.84	
Il − C'est	-3.42	0.27	-12.43	< 0.001	***
Aspect (perf)	-0.06	0.19	-0.33	0.74	
Alternation (alt)	1.02	0.19	5.46	< 0.001	***
(C'est − Cela):Aspect	-0.28	0.41	-0.67	0.50	
(Il − C'est):Aspect	-0.97	0.50	-1.97	0.05	*
		Spanish			
(Intercept)	0.41	0.25	1.64	0.10	
Esto − Este	5.18	0.62	8.34	< 0.001	***
Null − Esto	-4.32	0.59	-7.32	< 0.001	***
Aspect (perf)	0.32	0.28	1.16	0.25	
Alternation (alt)	1.16	0.30	3.91	< 0.001	***
		German			
(Intercept)	1.02	0.25	4.02	< 0.001	***
Dies − Das	0.11	0.55	0.20	0.84	
Es − Dies	-4.99	0.51	-9.87	< 0.001	***
Aspect (perf)	0.07	0.26	0.28	0.78	
Alternation (alt)	0.16	0.25	0.65	0.52	
		Italian			
(Intercept)	1.30	0.16	8.09	< 0.001	***
Ciò − Null	3.74	0.35	10.62	< 0.001	***
Questo − Ciò	-0.38	0.32	-1.19	0.24	
Aspect (alt)	0.89	0.27	3.26	0.001	**
Alternation (perf)	0.44	0.27	1.63	0.10	
(Ciò − Null):Altern	0.86	0.67	1.29	0.20	
(Questo − Ciò):Altern	0.81	0.63	1.28	0.20	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Appendix D: Verbs from ParCorFull and their spontaneity scores (cf. §5.3)

Here we present the list of verbs used with an entity (Table 23) or event (Table 24) antecedent in ParcorFull. The counts column comes from our extraction, while the causative-rate, anticausative-rate and spontaneity scores are taken from Samardžić (2014).

Table 23. ParcorFull verbs used with an entity antecedent.

Verb	Counts	Causative-rate	Anticausative-rate	Spontaneity-score
run	2	0.3	0.56	-0.64
grow	2	0.14	0.78	-1.68
open	2	0.54	0.14	1.33
shine	2	0.17	0.83	-1.61
split	1	0.24	0.05	1.54
tame	1	0.75	0.08	2.28
change	1	0.37	0.42	-0.14
burn	1	0.23	0.18	0.22
balance	1	0.18	0.05	1.34

Table 24. ParcorFull verbs used with an event antecedent.

Verb	Counts	Causative-rate	Anticausative-rate	Spontaneity-score
run	3	0.3	0.56	-0.64
fly	1	0.22	0.78	-1.27
drive	1	0.36	0.11	1.2
move	1	0.11	0.8	-1.97

References

Aarts, B. (2011). *Oxford modern English grammar*. Oxford University Press.

Akmajian, A. (1970). On deriving cleft sentences from pseudo-cleft sentences. *Linguistic Inquiry*, *1*(2), 149–168.

Allan, K. (2013). What is Common Ground?. In Capone A., Lo Piparo F., Carapezza M. (eds.), *Perspectives on Linguistic Pragmatics*, 285–310, Springer.

Almor, A. (1999). Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological Review*, *106*(4), 748–765.

Almor, A., & Nair, V.A. (2007). The Form of Referential Expressions in Discourse. *Language and Linguistics Compass*, *1*(1-2), 84–99.

Alonso-Ovalle, L., Férnandez-Solera, S., Frazier, L., & Clifton, C. J. (2002). Null vs. overt pronouns and the topic-focus articulation in Spanish. *Rivista di Linguistica*, *14*(2),151–170.

Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics*, *24*(1), 65–87.

Ariel, M. (1990). *Accessing Noun Phrase Antecedents*. Routledge.

Ariel, M. (1999). Cognitive universals and linguistic conventions: The case of resumptive pronouns. *Studies in Language*, *23*(2), 217–269.

Ariel, M. (2001). Accessibility theory: An overview. In T. Sanders, J. Schilperoord, & W. Spooren (Eds.), *Human Cognitive Processing*, Vol. 8. Text representation: Linguistic and psycholinguistic aspects, 29–87, John Benjamins.

Ariel, M. (2004). Accessibility Marking: Discourse Functions, Discourse Profiles, and Processing Cues. *Discourse Processes*, *37*(2), 91–116.

Arnold, J.E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes*, *3*, 137–162.

Arnold, J.E. (2010). How Speakers Refer: The Role of Accessibility. *Language and Linguistics Compass*, *4*, 187–203.

Arnold, J.E., & Griffin, Z.M. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of memory and language*, *56*(4), 521–536.

Asher, N. (1993). *Reference to Abstract Objects in Discourse*. Springer.

Azuma, H. (2008). A diachronic view of pronominal reference in English. In C. Johansson (ed.), *Second Workshop on Anaphora Resolution*, WAR II, volume 2, 1–9, Bergen.

Barr, D.J., Levy, R., Scheepers, C., & Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255–278.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015a). *Parsimonious mixed models*. https://arxiv.org/abs/1506.04967.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015b). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Baumann, P., Konieczny, L., & Hemforth, B. (2014). Conversational Implicatures in Anaphora Resolution: Alternative Constructions and Referring Expressions. In B. Hemforth, B. Mertins, & C. Fabricius-Hansen (eds.), *Psycholinguistic Approaches to Meaning and Understanding across Languages*, 197–212, Springer.

Bentzen, K., & Anderssen, M. (2019). The form and position of pronominal objects with non-nominal antecedents in Scandinavian and German. *The Journal of Comparative Germanic Linguistics*, *22*(2), 169–188.

Bergsma, S., & Yarowsky, D. (2011). NADA: A robust system for non-referential pronoun detection. In I. Hendrickx, S.L. Devi, A. Branco, & R. Mitkov (eds.), *Anaphora Processing and Applications: 8th Discourse Anaphora and Anaphor Resolution Colloquium (DAARC)*, Lecture Notes in Artificial Intelligence, 12–23. Springer.

Berruto, G. (2017). What is changing in Italian today? Phenomena of restandardization in syntax and morphology: an overview. In M. Cerruti, C. Crocco, & S. Marzo (eds.), *Towards a new standard*, 31–60. De Gruyter.

Betancort, M., Carreiras, M., & Sturt, P. (2009). The processing of subject and object relative clauses in Spanish: An eye-tracking study. *Quarterly Journal of Experimental Psychology*, *62*(10), 1915–1929.

Bevacqua, L., & Scheffler, T. (2020). Form Variation of Pronominal It-Clefts in Written English. A corpus study in Twitter and iWeb. *Linguistics Vanguard*, *6*(1).

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Longman.

Birkenes, M.B., & Sommer, F. (2015). The agreement of collective nouns in the history of Ancient Greek and German. In C. Gianollo, A. Jäger & D. Penka (eds.), *Language change at the syntax-semantics interface*, 183–222, De Gruyter.

Bock, K., Butterfield, S., Cutler, A., Cutting, J.C., Eberhard, K.M., & Humphreys, K.R. (2006). Number agreement in British and American English: Disagreeing to agree collectively. *Language*, *82*(1), 64–113.

Bock, K., Carreiras, M., & Meseguer, E. (2012). Number meaning and number grammar in English and Spanish. *Journal of Memory and Language*, *66*(1), 17–37.

Bock, K., Eberhard, K.M., & Cutting, J.C. (2004). Producing number agreement: How pronouns equal verbs. *Journal of Memory and Language*, *51*, 251–278.

Bojar, O., Chatterjee, R., Federmann, C., Graham, Y., Haddow, B., Huang, S., Huck, M., Koehn, P., Liu, Q., Logacheva, V., Monz, C., Negri, M., Post, M., Rubino, R., Specia, L.,

& Turchi, M. (2017). Findings of the 2017 conference on machine translation (wmt17). In *Proceedings of the Second Conference on Machine Translation*, Volume 2: Shared Task Papers, 169–214, Copenhagen, Denmark, Association for Computational Linguistics.

Bonnefon, J.-F., Feeney, A., & Villejoubert, G. (2009), When some is actually all: Scalar inferences in fact-threatening contexts. *Cognition*, *112*(2), 249–258.

Botley, S.P., & McEnery, T. (eds.) (2000). *Corpus-based and computational approaches to discourse anaphora*. John Benjamins.

Bouma, G., Øvrelid, L., & Kuhn, J. (2010). Towards a large parallel corpus of clefts. In *Proceedings of the Conference on Language Resources and Evaluation (LREC)*, Valetta, Malta.

Boyd, A., Gegg-Harrison, W., & Byron, D.K. (2005). Identifying non-referential it: a machine learning approach incorporating linguistically motivated patterns. In *Proceedings of the ACL Workshop on Feature Engineering for Machine Learning in Natural Language Processing*, 40–47, Ann Arbor, Michigan, Association for Computational Linguistics.

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*(3), 434–463.

Brennan, S.E. (1995). Centering attention in discourse. *Language and Cognitive Processes*, *10*(2), 137–167.

Brennan, S.E., Friedman, M.W., & Pollard, C.J. (1987). A centering approach to pronouns. In *Proceedings of the 25th annual meeting on Association for Computational Linguistics (ACL '87)*, 155–162, Association for Computational Linguistics.

Brown-Schmidt, S., & Duff, M. C. (2016). Memory and Common Ground Processes in Language Use. *Topics in cognitive science*, *8*(4), 722–736.

Brown-Schmidt, S., Byron, D.K., & Tanenhaus, M.K. (2005). Beyond salience: Interpretation of personal and demonstrative pronouns. *Journal of Memory and Language*, *53*(2), 292–313.

Büring, D. (2005). *Binding Theory*. Cambridge University Press.

Byron, D.K. (2002). Resolving pronominal reference to abstract entities. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, ACL 2002, 80–87, Philadelphia, Association for Computational Linguistics.

Cambridge Dictionary (2019). *Grammar*. https://dictionary.cambridge.org/grammar/british-grammar/pronouns-personal-i-me-you-him-it-they-etc (Accessed 4 May 2020).

Cardinaletti, A. (2004). Towards a cartography of subject positions. In L. Rizzi (ed.), *The structure of CP and IP*. The cartography of syntactic structures, vol. 2, 115–165, Oxford University Press.

Carlson, G. (2003). Reference. In L.R. Horn & G. Ward (eds.), *The Handbook of Pragmatics*. Wiley-Blackwell.

Carminati, M.N. (2002). *The processing of Italian subject pronouns*. Ph.D. thesis, University of Massachusetts Amherst.

Chafe, W. (1976). Giveness, contrastiveness, definiteness, subjects, topics and point of view. In C.N. Li (ed.), *Subject and Topic*, 25–56, Academic Press.

Chafe, W. (1987). Cognitive constraints on information flow. In S. Tomlin (ed.), *Coherence and Grounding in Discourse: Outcome of a Symposium*, 21–51, John Benjamins.

Cheshire, J., Adger, D., & Fox, S. (2013). Relative who and the actuation problem. *Lingua*, *126*, 51–77.

Chomsky, N. (1981). *Lectures on government and binding*. Foris Publications.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge.

Çokal, D., Sturt, P., & Ferreira, F. (2014). Deixis: This and That in written narrative discourse. *Discourse Processes*, *51*(3), 201–229.

Çokal, D., Sturt, P., & Ferreira, F. (2018). Processing of It and This in Written Narrative Discourse. *Discourse Processes*, *55*(3), 272-289.

Collins, P. C. (1991). *Cleft and Pseudo-Cleft Constructions in English*. Routledge.

Colonna, S., Schimke, S., & Hemforth, B. (2010). Le role de la structure informationnelle dans l'interpretation d'une anaphore pronominale interphrastique. In *Proceedings of the CMLP 2010 – 2ème congrès mondial de linguistique français*, 1489–1499, New Orleans.

Colonna, S., Schimke, S., & Hemforth, B. (2012). Information structure effects on anaphora resolution in German and French: A cross-linguistic study of pronoun resolution. *Linguistics*, *50*(5), 991–1013.

Comrie, B. (1997). Pragmatic binding: Demonstratives as anaphors in Dutch. In *Proceedings of the Twenty-Third Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on Pragmatics and Grammatical Structure*, 50–61, Berkeley.

Corbett, G.G. (1979). The agreement hierarchy. *Journal of Linguistics*, *15*(2), 203–224.

Corbett, G.G. (2000). *Number*. Cambridge University Press.

Cornish, F. (2015). *Anaphoric Relations in English and French: A Discourse Perspective*. Taylor & Francis.

Cowles, H.W., & Garnham, A. (2011). Noun-Phrase Anaphor Resolution: Antecedent Focus, Semantic Overlap, and the Informational Load Hypothesis. In E. Gibson & N.J. Pearlmutter (eds.), *The Processing and Acquisition of Reference*, 297–322, MIT Press.

Cowles, H.W., Walenski, M., & Kluender, R. (2007). Linguistic and cognitive prominence in anaphor resolution: topic, contrastive focus and pronouns. *Topoi*, *26*, 3–18.

Crawley, R. A., Stevenson, R. J., & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research*, *19*(4), 245–264.

Cummings, L. (2018). *Working with English grammar: An introduction*. Cambridge University Press.

Davies, M. (2018). *The 14 Billion Word iWeb Corpus*. https://www.english-corpora.org/iweb/ (Accessed 10 December 2020).

de Carvalho Maia, J., Vernice, M., Gelormini Lezama, C., Lima, M. L., & Almor, A. (2017). Co-referential processing of pronouns and repeated names in Italian. *Journal of Psycholinguistic Research*, *46*(2), 497–506.

de la Fuente, I., & Hemforth, B. (2013). Effects of clefting and left-dislocation on subject and object pronoun resolution in Spanish. In J. Cabrelli Amaro, G. Lord, A. de Prada Pérez, & J. Aaron (eds.), *Selected proceedings of the hispanic linguistics symposium 2012*, 27–45, Cascadilla Press.

Delin, J. (1992). Properties of It-Cleft Presupposition. *Journal of Semantics*, *9*(4), 289–306.

Depraetere, I. (2003). On verbal concord with collective nouns in British English. *English Language and Linguistics*, *7*(1), 85–127.

Depraetere, I., & Langford, C. (2019). *Advanced English grammar: A linguistic approach*. Bloomsbury.

Dipper, S. & Zinsmeister, H. (2010). Towards a standard for annotating abstract anaphora. In *Proceedings of the LREC Workshop on Language Resource and Language Technology Standards state of the art, emerging needs, and future developments*, LREC10-W4, 54–59, Valletta, Malta, European Language Resources Association (ELRA).

Dipper, S., Rieger, C., Seiss, M., & Zinsmeister, H. (2011). Abstract anaphors in German and English. In I. Hendrickx, S.L. Devi, A. Branco, & R. Mitkov (eds.), *Anaphora Processing and Applications. 8th Discourse Anaphora and Anaphor Resolution Colloquium*, DAARC 2011, 96–107, Springer.

Dudenredaktion (2009). *Duden. Die Grammatik: Unentbehrlich für richtiges Deutsch*. Dudenverlag.

Eastwood, J. (1994). *Oxford guide to English grammar*. Oxford University Press.

Eberhard, K.M. (1999). The accessibility of conceptual number to the processes of subject–verb agreement in English. *Journal of Memory and Language*, *41*(4), 560–578.

Eckert, M., & Strube, M. (2000). Dialogue acts, synchronising units and anaphora resolution. *Journal of Semantics*, *17*(1), 51–89.

Eisenstein, J. (2013a). Phonological Factors in Social Media Writing. In *Proceedings of the Workshop on Language in Social Media (LASM 2013)*, 11–19.

Eisenstein, J. (2013b). What to do about bad language on the internet. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 359–369.

Embick, D. (2004). Unaccusative syntax and verbal alternations. In A. Alexiadou, E. Anagnostopoulou, & M. Everaert (eds.), *The Unaccusativity Puzzle: Explorations of the Syntax-Lexicon Interface*, 137–158, Oxford University Press.

Evans, E. (2001). Applying machine learning toward an automatic classification of it. *Literary and Linguistic Computing*, *16*(1), 45–57.

Fedele, E., & Kaiser, E. (2014). Looking back and looking forward: Anaphora and cataphora in Italian. *University of Pennsylvania Working Papers in Linguistics*, *20*(1), 81–90.

Fernandes, E.G., Luegi, P., Correa Soares, E., de la Fuente, I., & Hemforth, B. (2018). Adaptation in pronoun resolution: Evidence from Brazilian and European Portuguese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(12), 1986–2008.

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*(4), 296–340.

Ferretti, T.R., Kutas, M., & McRae, K. (2007). Verb aspect and the activation of event knowledge. *Journal of experimental psychology. Learning, memory, and cognition*, *3*(1), 182–196.

Ferretti, T.R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*, *44*(4), 516–547.

Ferretti, T.R., Rohde, H., Kehler, A., & Crutchley, M. (2009). Verb aspect, event structure, and coreferential processing. *Journal of memory and language*, *61*(2), 191–205.

Filiaci, F. (2010). Null and overt subject biases in Spanish and Italian: a cross-linguistic comparison. In C. Borgonovo, M. Español-Echevarria, & P. Prevost (eds.), *Selected Proceedings of the 12th Hispanic Linguistics Symposium*, 171–182, Cascadilla Press.

Filiaci, F., Sorace, A., & Carreiras, M. (2014): Anaphoric biases of null and overt subjects in Italian and Spanish: a cross-linguistic comparison. *Language, Cognition and Neuroscience*, *29*(7), 825–843.

Franke, M., & Jäger, G. (2016). Probabilistic pragmatics, or why Bayes' rule is probably important for pragmatics. *Zeitschrift für Sprachwissenschaft*, *35*(1), 3–44.

Frascarelli, M. (2007). Subjects, topics and the interpretation of referential pro. *Natural Language & Linguistic Theory*, *25*(4), 691–734.

Fries, U. (1988). The crew have abandoned the ship: Concord with collective nouns revisited. *AAA: Arbeiten aus Anglistik und Amerikanistik*, *13*(2), 99–104.

Fukumura, K., & van Gompel, R.P.G. (2010). Choosing anaphoric expressions: Do people take into account likelihood of reference?. *Journal of Memory and Language*, *62*(1), 52–66.

Fukumura, K., & van Gompel, R.P.G. (2011). The effect of animacy on the choice of referring expression. *Language and Cognitive Processes*, *26*(10), 1472–1504.

Fukumura, K., & van Gompel, R.P.G. (2012). Producing Pronouns and Definite Noun Phrases: Do Speakers Use the Addressee's Discourse Model?. *Cognitive Science*, *36*(7), 1289–1311.

Geber, D. (2006), Processing subject pronouns in relation to non-canonical (quirky) constructions. *Cahiers linguistiques d'Ottawa*, *34*, 47–61.

Gelormini Lezama, C. (2018). Exploring the repeated name penalty and the overt pronoun penalty in Spanish. *Journal of Psycholinguistic Research*, *47*(2), 377–389.

Gelormini-Lezama, C. & Almor, A. (2011) Repeated names, overt pronouns and null pronouns in Spanish. *Language and Cognitive Processes*, *26*(3), 437-454.

Gelormini-Lezama, C. & Almor, A. (2014). Singular and plural references in Spanish. *Journal of Psycholinguistic Research*, *43*(3), 299-313.

Gernsbacher, M.A. (1990). *Language comprehension as structure building*. Lawrence Erlbaum Associates.

Geurts, B. (2019). Communication as commitment sharing: speech acts, implicatures, common ground. *Theoretical Linguistics*, *45*(1-2), 1–30.

Gibson, E., Piantadosi, S., & Fedorenko, K. (2011). Using Mechanical Turk to obtain and analyze English acceptability judgments. *Language and Linguistics Compass*, *5*(8), 509–524.

Givón, T. (1983). *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. John Benjamin.

Givón, T. (1985). Iconicity, isomorphism, and non-arbitrary coding in syntax. In J. Haiman (ed.), *Iconicity in Syntax*, 187–220, John Benjamin.

Givón, T. (1989). *Mind, Code and Context. Essays in Pragmatics*. Lawrence Erlbaum Associates.

Gordon, P.C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, *17*(3), 311–347.

Gordon, P.C., Hendrick, R., Ledoux, K., & Yang, C.L. (1999). Processing of reference and the structure of language: An analysis of complex noun phrases. *Language and Cognitive Processes*, *14*(4), 353–379.

Grammar Book (2019). https://www.grammarbook.com/grammar/pronoun.asp (Accessed 10 December 2020).

Grass, T. (2002). *Quoi! vous voulez traduire «goethe»?*. Peter Lang.

Greene, S.B., McKoon, G., & Ratcliff, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*, 266–283.

Grevisse, M., & Goosse, A. (2008). *Le bon usage: Grammaire française*, 14th edition. Duculot.

Grice, P. (1957). Meaning. *The Philosophical Review*, *66*(3), 377–388.

Grice, P. (1989). *Studies in the way of words*. Harvard University Press.

Grodner, D.J., Klein, N.M., Carbary, K.M., & Tanenhaus, M.K. (2010). "Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*(1), 42–55.

Grosz, B.J., Joshi, A.K., & Weinstein, S. (1983). Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st annual meeting on Association for Computational Linguistics (ACL '83)*, 44–50, Association for Computational Linguistics.

Grosz, B.J., Joshi, A.K., & Weinstein, S. (1995). Centering: A Framework for Modeling the Local Coherence of Discourse. *Computational Linguistics*, *21*(2), 203–225.

Guillou, L. (2016). *Incorporating Pronoun Function into Statistical Machine Translation*. Ph.D. thesis, University of Edinburgh.

Guillou, L., Hardmeier, C., Smith, A., Tiedemann, J., & Webber, B. (2014). ParCor 1.0: A parallel pronoun-coreference corpus to support statistical MT. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*, LREC 2014, 3191–3198, Reykjavik, European Language Resources Association (ELRA).

Gundel, J.K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, *69*(2), 274–307.

Gundel, J.K., Hedberg, N., & Zacharski, R. (2005). Pronouns without np antecedents: How do we know when a pronoun is referential? In A. Branco, T. McEnery, & R. Mitkov (eds.), *Anaphora Processing: Linguistic, Cognitive and Computational Modelling*, 351–364, John Benjamins.

Haiman, J. (1985). *Natural Syntax*. Cambridge University Press.

Halliday, M.A.K., & Hasan, R. (1976). *Cohesion in English*. Longman.

Hardmeier, C., Bevacqua, L., Loáiciga, S., & Rohde, H. (2018). *Forms of anaphoric reference to organisational named entities: Hoping to widen appeal, they diversified*. In Proceedings of the seventh named entities workshop, 36–40, Association for Computational Linguistics.

Hardmeier, C., Tiedemann, J., Nakov, P., Stymne, S., & Versely, Y. (2016). DiscoMT 2015 Shared Task on Pronoun Translation. LINDAT/CLARIN digital library at Institute of Formal and Applied Linguistics, Charles University in Prague.

Haspelmath, M. (1993). More on the typology of inchoative/causative verb alternations. In B. Comrie & M. Polinsky (eds.), Causatives and transitivity, 87–120, John Benjamins.

Hawksey, M. (2016). *TAGS: Twitter archiving Google spreadsheet*, v. 6.1. https://tags.hawksey.info/ (Accessed 9 December 2020).

Heath, M. (2018). Orthography in social media: Pragmatic and prosodic interpretations of caps lock. *Proceedings of the Linguistic Society of America*, *3*(55), 1–13.

Hedberg, N., Gundel, J.K., & Zacharski, R. (2007). Directly and indirectly anaphoric demonstrative and personal pronouns in newspaper articles. In *Proceedings of the Discourse Anaphora and Anaphora Resolution Colloquium*, DAARC 2007, Lagos (PT).

Heinzerling, B., Moosavi, N.S., & Strube, M. (2017). Revisiting selectional preferences for coreference resolution. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 1332–1339, Copenhagen, Denmark, Association for Computational Linguistics.

Hemforth, B., Konieczny, L., Scheepers, C., Colonna, S., Schimke, S., Baumann, P., & Pynte, J. (2010). Language specific preferences in anaphor resolution: Exposure or gricean maxims? In *32nd annual conference of the cognitive science society*, 2218–2223.

Heylighen, F., & Dewaele, J.-M. (1999). *Formality of Language: Definition, measurement and behavioral determinants*. Internal Report, Center "Leo Apostel", Free University of Brussels.

Heylighen, F., & Dewaele, J.-M. (2002). Variation in the contextuality of language: An empirical measure. *Foundations of Science*, *7*, 293–340.

Hilpert, M. (2014). *Construction grammar and its application to English*. Edinburgh University Press.

Hobbs, J.R. (1979). Coherence and coreference. *Cognitive Science*, *3*(1), 67–90.

Hope, A.C.A. (1968). A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society. Series B (Methodological)*, *30*(3), 582–598.

Hu, Y., Talamadupula, K., & Kambhampati, S. (2013). Dude, srsly?: The Surprisingly Formal Nature of Twitter's Language. In *Proceedings of the International Conference on Weblogs and Social Media*.

Huang, Y. (1991). A neo-Gricean pragmatic theory of anaphora. *Journal of Linguistics*, *27*(2), 301–335.

Huang, Y. (2000a). *Anaphora. A cross-linguistic Approach*. Oxford University Press.

Huang, Y. (2000b). Discourse anaphora: Four theoretical models. *Journal of Pragmatics*, *32*(2), 151–176.

Huang, Y.T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: insight into the semantics-pragmatics interface. *Cognitive psychology*, *58*(3), 376–415.

Huddleston, R. & Pullum, J.K. (2008). *The Cambridge grammar of the English language*. Cambridge University Press.

Humphreys, K.R., & Bock, K. (2005). Notional number agreement in English. *Psychonomic Bulletin Review*, *12*, 689–695.

Hundt, M. (2006). The committee has/have decided… On concord patterns with collective nouns in inner- and outer-circle varieties of English. *Journal of English Linguistics*, *34*(3), 206–232.

Hundt, M. (2009). Concord with collective nouns in Australian and New Zealand English. In Peters, P., Collins, P., & Smith, A. (eds.), *Comparative studies in Australian and New Zealand English: Grammar and beyond*, 207–224, John Benjamins.

Hwang, H. (2020). Avoidance of gender-ambiguous pronouns as a consequence of ambiguity-avoidance strategy. *Discourse Processes*, https://doi.org/10.1080/0163853X.2020.1844965 (Accessed 9 December 2020).

Istat (2018). *Natalità e fecondità della popolazione residente*. Istituto Nazionale di Statistica (ISTAT).

Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*. MIT Press.

Janner, M.C. (2018). *Sguardi linguistici sulla marca*. Peter Lang.

Jespersen, O. (1927). *Modern English grammar on historical principles*, Part III: Syntax, vol. 2. Allen and Unwin.

Kaiser, E. (2011). Focusing on pronouns: Consequences of subjecthood, pronominalisation, and contrastive focus. *Language and Cognitive Processes*, *26*, 1625–1666.

Kaiser, E., & Trueswell, J.C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, *94*, 113–147.

Kameyama, M. (1999). Stressed and Unstressed Pronouns: Complementary Preferences. In P. Bosch & R. van der Sandt (eds.), *Focus: Linguistic, Cognitive, and Computational Perspectives*, 306–321, Cambridge University Press.

Kehler, A. (2002). *Coherence, reference and the theory of grammar*. CSLI Publications.

Kehler, A., & Rohde, H. (2018). Prominence and coherence in a Bayesian theory of pronoun interpretation. J*ournal of Pragmatics*, *154*, 63–78.

Kehler, A., Kertz, L., Rohde, H., & Elman, J. (2008). Coherence and Coreference Revisited. *Journal of Semantics*, *25*(1), 1–44.

Kjellmer, G. (1997). On contraction in modern English. *Studia Neophilologica*, *69*(2), 155–186.

Klin, C.M., Weingartner, K.M., Guzmán, A.E., & Levine, W.H. (2004). Readers' sensitivity to linguistic cues in narratives: How salience influences anaphor resolution. *Memory & Cognition*, *32*, 511–522.

Koch, P., & Oesterreicher, W. (1985). Sprache der Nähe – Sprache der Distanz. *Romanistisches Jahrbuch*, *36*, 15–43.

Kolhatkar, V., Roussel, A., Dipper, S., & Zinsmeister, H. (2018). Anaphora with non-nominal antecedents in computational linguistics: a Survey. *Computational Linguistics*, *44*(3), 547–612.

Kolhatkar, V., Zinsmeister, H., & Hirst, G. (2013). Interpreting anaphoric shell nouns using antecedents of cataphoric shell nouns as training data. In *Proceedings of the 2013*

Conference on Empirical Methods in Natural Language Processing, 300–310, Seattle, Association for Computational Linguistics.

Kreiner, H., Garrod, S., & Sturt, P. (2013). Number agreement in sentence comprehension: The relationship between grammatical and conceptual factors. *Language and Cognitive Processes*, *28*(6), 829–874.

Lakaw, A. (2017). Diachronic shifts in agreement patterns of collective nouns in 19th-century American and British English. *Studies in Variation, Contacts and Change in English*, *18*.

Lambrecht, K. (1994). *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge University Press.

Lambrecht, K. (2000). When subjects behave like objects: An analysis of the merging of S and O in Sentence-Focus Constructions across languages. *Studies in Language*, *24*(3), 611–682.

Lambrecht, K. (2001). A framework for the analysis of cleft constructions. *Linguistics*, *39*(3), 463–516.

Lammert, M. (2010). *Sémantique et cognition. Les noms collecetifs*. Droz.

Lapshinova-Koltunski, E., & Hardmeier, C. (2018). *Coreference corpus annotation guidelines*. https://lindat.mff.cuni.cz/repository/xmlui/bitstream/handle/11372/LRT-2614/Guidelines.pdf (Accessed 10 December 2020).

Lapshinova-Koltunski, E., Hardmeier, C., & Krielke, P. (2018). ParCorFull: a parallel corpus annotated with full coreference. In *Proceedings of 11th Language Resources and Evaluation Conference*, 423–428, Miyazaki, Japan, European Language Resources Association (ELRA).

Lee, K., He, L., Lewis, M., & Zettlemoyer, L. (2017). End-to-end neural coreference resolution. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language*

Processing, 188–197, Copenhagen, Denmark, Association for Computational Linguistics.

Lee, T., Lutz, A., & Choi, J.D. (2016). QA-It: classifying non-referential it for question answer pairs. In *Proceedings of the ACL 2016 Student Research Workshop*, 132–137, Berlin, Germany, Association for Computational Linguistics.

Levin, B. (1993). *English verb classes and alternations: a preliminary investigation*. The University of Chicago Press.

Levin, M. (2001). *Agreement with collective nouns in English*. Ph.D. thesis, Lund University.

Levinson, S.C. (1987). Pragmatics and the grammar of anaphora: a partial pragmatic reduction of Binding and Control phenomena. *Journal of Linguistics*, *23*, 379–434.

Levinson, S.C. (1991). Pragmatic Reduction of the Binding Conditions Revisited. *Journal of Linguistics*, *27*(1), 107–161.

Levinson, S.C. (1995). Three levels of meaning. In F. Palmer (ed.), *Grammar and meaning*, 90–115, Cambridge University Press.

Levinson, S.C. (2000). *Presumptive Meanings. The Theory of Generalized Conversational Implicature*. MIT Press.

Loáiciga, S., Bevacqua, L., Rohde, H., & Hardmeier, C. (2018). Event versus entity co-reference: Effects of context and form of referring expression. In *Proceedings of the First Workshop on Computational Models of Reference, Anaphora and Coreference*, 97–103, New Orleans, Louisiana, Association for Computational Linguistics.

Loáiciga, S., Guillou, L., & Hardmeier, C. (2017). What is it? disambiguating the different readings of the pronoun "it". In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 1325–1331, Copenhagen, Denmark, Association for Computational Linguistics.

Loáiciga, S., Hardmeier, C., & Sayeed, A. (2020). Exploiting cross-lingual hints to discover event pronouns. In *Proceedings of the 12th Language Resources and Evaluation Conference*, 99–103, Marseille, France, European Language Resources Association.

MacKay, D.G., & Fulkerson, D.C. (1979). On the comprehension and production of pronouns. *Journal of Verbal Learning & Verbal Behavior*, *18*(6), 661–673.

Maier, G. (2013). As the case may be: A corpus-based approach to pronoun case variation in subject predicative complements in British and American English. In M. Huber & J. Mukherjee (eds.), *Corpus Linguistics and Variation in English: Focus on Non-Native Englishes*, Vol. 13, VARIENG.

Maier, G. (2014). The case of focus. In K. Davidse, C. Gentens, L. Ghesquière & L. Vandelanotte (eds.), *Corpus Interrogation and Grammatical Patterns*, 173–205, John Benjamins.

Marasovic, A., Born, L., Opitz, J., & Frank, A. (2017). A mention-ranking model for abstract anaphora resolution. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 221–232, Copenhagen, Denmark, Association for Computational Linguistics.

Mayol, L. (2018). Asymmetries between interpretation and production in Catalan pronouns. *Dialogue and Discourse*, *9*(2), 1–34.

McAteer, E. (1992). Typeface Emphasis and information focus in written language. *Applied Cognitive Psychology*, *6*, 345–359.

Moens M., & Steedman, M. (1988). Temporal ontology and temporal reference. *Computational Linguistics*, *14*(2), 15–28.

Müller, C. (2007). Resolving It, This, and That in unrestricted multi-party dialog. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, ACL07, 816–823, Prague, Czech Republic, Association for Computational Linguistics.

Munro, R., Bethard, S., Kuperman, V., Lai, V.T., Melnick, R., Potts, C., Schnoebelen, T., & Tily, H. (2010). Crowd-sourcing and language studies: the new generation of linguistic data. In *Proceedings of the NAACLHLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, 122–130, Association for Computational Linguistics.

Navarretta, C. (2007). A contrastive analysis of abstract anaphora in Danish, English and Italian. In A. Branco, T. McEnery, R. Mitkov, & F. Silva (eds.), *Proceedings of the Discourse Anaphora and Anaphora Resolution Colloquium*, DAARC 2007, 103–109, Centro de Linguística da Universida de do Porto.

Nedoluzhko A., & Lapshinova-Koltunski, E. (2016). Abstract coreference in a multilingual perspective: a view on Czech and German. In *Proceedings of the Workshop on Coreference Resolution Beyond OntoNotes*, COR-BON 2016, 47–52, Ann Arbor, Michigan, Association for Computational Linguistics.

Nelson, G., Wallis, S., & Aarts, B. (1998). *International Corpus of English – Great Britain (ICE-GB)*, Release 2. https://www.ucl.ac.uk/english-usage/projects/ice-gb/ (Accessed 10 December 2020).

Noveck, I.A., & Posada, A. (2003). Characterizing the time course of an implicature: an evoked potentials study. *Brain and Language*, *85*(2), 203–210.

Palmer, M., Gildea, D., & Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, *31*(1), 71–106.

Passonneau, R.J. (1989). Getting at discourse referents. In *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, 51–59, Vancouver, Association for Computational Linguistics.

Patten, A.L. (2010). *Cleft sentences, construction grammar and grammaticalization*. Ph.D. thesis, University of Edinburgh.

Patten, A.L. (2012). *The English it-Cleft. A Constructional Account and a Diachronic Investigation*. De Gruyter Mouton.

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics*, *27*(1), 89–110.

Perera, J., & Bel, A. (2011), Propietats pragmàtiques i gramaticals en el desenvolupament de la coreferència pronominal. *Zeitschrift für Katalanistik*, *24*, 183–211.

Perlmutter, D.M. (1972). A note on syntactic and semantic number in English. *Linguistic Inquiry*, *3*(2), 243–246.

Poesio, M. (2016). Linguistic and cognitive evidence about anaphora. In M. Poesio, R. Stuckardt & Y. Versley (eds.), *Anaphora resolution: Algorithms, resources, and applications*, 23–54, Springer.

Poesio, M., Stevenson, R., Di Eugenio, B., & Hitzeman, J. (2004). Centering: A parametric theory and its instantiations. *Computational Linguistics*, *30*(3), 309–363.

Poesio, M., Stuckardt, R., & Versley, Y. (2015). Challenges and directions of further research. In M. Poesio, R. Stuckardt, & Y. Versley (eds.), *Anaphora Resolution: Algorithms, Resources and Application*, 487–500, Springer-Verlag.

Poesio, M., Stuckardt, R., & Versley, Y. (eds.) (2016). *Anaphora Resolution. Algorithms, Resources, and Applications*. Springer-Verlag.

Pollard, C., & Sag, I.A. (1992). Anaphors in English and the Scope of Binding Theory. *Linguistics Inquiry*, *23*(2), 261–303.

Pradhan, S., & Xue, N. (2009). OntoNotes: The 90% solution. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Companion Volume: Tutorial Abstracts, 11–12, Boulder, Colorado, Association for Computational Linguistics.

Pradhan, S., Moschitti, A., Xue, N., Ng, H.T., Björkelund, A., Uryupina, O., Zhang, Y., & Zhong, Z. (2013). Towards robust linguistic analysis using OntoNotes. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*, 143–152, Sofia, Bulgaria, Association for Computational Linguistics.

Prince, E. F. (1978). A Comparison of Wh-Clefts and it-Clefts in Discourse. *Language*, *54*(4), 883–906.

Pynte, J., & Colonna, S. (2000). Decoupling syntactic parsing from visual inspection: the case of relative clause attachment in French. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (eds.), *Reading as a perceptual process*, 529–547, Elsevier.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. Longman.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Real Academia Española (RAE), (2010). *Nueva gramática de la lengua española. Manual*. Real Academia Española y Asociación de Academias de la Lengua Española.

Reinhart, T. (1983). *Anaphora and Semantic Interpretation*. University of Chicago Press.

Renzi, L. (1994). Egli - lui - il - lo. In T. De Mauro (ed.), *Come parlano gli italiani*, 247–250, La Nuova Italia.

Renzi, L. (2012). *Come cambia la lingua. L'italiano in movimento*. Il Mulino.

Rizzi, L. (1997). A parametric approach to comparative syntax properties. In L. Haegeman (ed.), *The new comparative syntax*, 268–275, Longman.

Rohde, H. (2008). *Coherence-Driven Effects in Sentence and Discourse Processing*. Ph.D. thesis, University of California, San Diego.

Rohde, H. (2019). Pronouns. In C. Cummins & N. Katsos (eds.), *Oxford Handbook of Experimental Semantics and Pragmatics*, 452–473, Oxford University Press.

Rohde, H., & Kehler, A. (2014). Grammatical and information-structural influences on pro-noun production. *Language, Cognition and Neuroscience*, *29*(8), 912–927.

Rosa, E. C., & Arnold, J. E. (2017). Predictability affects production: Thematic roles can affect reference form selection. *Journal of Memory and Language*, *94*, 43–60.

Samardžić, T. (2014). *Dynamics, causation, duration in the predicate-argument structure of verbs: A computational approach based on parallel corpora*. PhD thesis, University of Geneva, Switzerland.

Schäfer, F. (2009). The causative alternation. *Language and Linguistics Compass*, *3*(2), 641–681.

Scheffler, T. (2017). Conversations on Twitter. In D. Fišer & M. Beißwenger (eds.), *Investigating Computer-Mediated Communication: Corpus-Based Approaches to Language in the Digital World*, 124–144, Ljubljana University Press.

Schielzeth, H., Dingemanse, N.J., Nakagawa, S., Westneat, D.F., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N.A., Garamszegi, L.Z., & Araya-Ajoy, Y.G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, *11*(9), 1141–1152.

Schweppe, J. (2013). Distance effects in number agreement. *Discourse Processes*, *50*, 531–556.

Scott, K. (2015). The pragmatics of hashtags: Inference and conversational style on Twitter. *Journal of Pragmatics*, *81*, 8–20.

Serianni, L. (1997). *Italiano. Grammatica, Sintassi, Dubbi*. Garzanti.

Serratrice, L. (2007). Cross-linguistic influence in the interpretation of anaphoric and cataphoric pronouns in English–Italian bilingual children. *Bilingualism: Language and Cognition*, *10*, 225–238.

Slabakova, R. (2010). Scalar implicatures in second language acquisition. *Lingua*, *120*, 2444–2462.

Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, *22*(3), 339–368.

Spector, B. (2014). Global positive polarity items and obligatory exhaustivity. *Semantics & Pragmatics*, *7*(11), 1–61.

Sperber, D., & Wilson, D. (1982). Mutual knowledge and relevance in theories of comprehension. In N.V. Smith (ed.), *Mutual Knowledge*, 61–131, Academic Press.

Sperber, D., & Wilson, D. (1995). *Relevance. Communication & Cognition* (2nd edition). Blackwell.

Sperber, D., & Wilson, D. (2002). Pragmatics, Modularity and Mind-reading. *Mind and Language*, *17*(1–2), 3–23.

Storrer, A. (2013). Sprachstil und Sprachvariation in sozialen Netzwerken. In B. Frank-Job, A. Mehler, & T. Sutter (eds.), *Die Dynamik sozialer und sprachlicher Netzwerke. Konzepte, Methoden und empirische Untersuchungen an Beispielen des WWW*, VS Verlag für Sozialwissenschaften.

Taboada, M. (2008). Reference, Centers, and Transitions in Spoken Spanish. In J. Gundel & N. Hedberg (eds.), *Reference: Interdisciplinary Perspectives*, 176–215, Oxford University Press.

Tagliamonte, S. A., & Denis, D. (2008). Linguistic ruin? LOL! Instant messaging and teen language. *American Speech*, *83*(1), 3–34.

Tily, H., & Piantadosi, S. (2009). Refer efficiently: Use less informative expressions for more predictable meanings. In *Proceedings of the workshop on the production of referring expressions: Bridging the gap between computational and empirical approaches to reference*, Cognitive Science Society.

Tristram, A. (2012). *Variation and change in verbal agreement with collective nouns in French*. Ph.D. thesis, University of Cambridge.

Tristram, A. (2014). Diachronic change in verbal agreement patterns with majorité. *Transactions of the Philological Society*, *112*(3), 344–366.

Tristram, A. (2015). L'accord sujet-verbe en français contemporain: Une étude decas – la/une foule. *Revue Romane*, *50*(2), 191–221.

Tristram, A., & Ayres-Bennett, W. (2012). From negation to agreement: revisiting the problem of sources for socio-historical linguistics. *Neuphilologische Mitteilungen*, *113*, 365–393.

Tutin, A. (2002). A corpus-based study of pronominal anaphoric expressions in French. In *Proceedings of the Discourse Anaphora and Anaphora Resolution Colloquium*, DAARC 2002, Lisbon, 265–277.

Ueno, M., & Kehler, A. (2016). Grammatical and pragmatic factors in Japanese pronoun interpretation. *Linguistics*, *54*, 1165–1221.

Uryupina, O., Artstein, R., Bristot, A., Cavicchio, F., Delogu, F., Rodriguez, K., & Poesio, M. (2020). Annotating a broad range of anaphoric phenomena, in multiple genres: the ARRAU corpus. *Natural Language Engineering*, *26*, 95–128.

Vallduví, E. (1993). *Information packaging: A survey*. Technical Report No. HCRC/RP-4, Centre for Cognitive Science and Human Communication Research, University of Edinburgh.

van Gompel, R.P.G., & Majid, A. (2004). Antecedent frequency effects during the processing of pronouns. *Cognition*, *90*(3), 255–264.

Vande Casteele, A. (2012). A semantic description of company names in Spanish business-related newspaper articles. In R. Boerrigter & H. Nijboer (eds.), *Names as language and capital*, 93–103, Meertens Instituut.

Versley, Y., & Björkelund, A. (2015). Off-the-shelf tools. In M. Poesio, R. Stuckardt, & Y. Versley (eds.), *Anaphora Resolution: Algorithms, Resources and Application*, 237–266, Springer-Verlag.

Vieira, R., Salmon-Alt, S., Gasperin, C., Schang, E., & Othero, G. (2005). Coreference and anaphoric relations of demonstrative noun phrases in multilingual corpus. In A. Branco, T. McEnery, & R. Mitkov (eds.), *Anaphora Processing: linguistic, cognitive and computational modeling*, 385–403, John Benjamins.

Von Heusinger, K. (1997). Salience and definiteness. *The Prague Bulletin of Mathematical Linguistics*, *67*, 5–23.

Von Heusinger, K., & Schumacher, P.B. (2019), Discourse prominence: Definition and application. *Journal of Pragmatics*, *154*, 117–127

Webber, B. (1986). Findings of the 2016 WMT shared task on cross-lingual pronoun prediction. In *Theoretical Issues in Natural Language Processing*, TINLAP-3, 158–163, Las Cruces, New Mexico, Association for Computational Linguistics.

Webber, B. (1990). *Structure and ostension in the interpretation of discourse deixis*. Technical Report MS-CIS-90-58, University of Pennsylvania, Department of Computer and Information Science.

Wikipedia contributors (2019). *Manual of style* — Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/ (Accessed 4 May 2020).

Wilson, D., & Sperber, D. (2004). Relevance Theory. In L.R. Horn & G. Ward (eds.), *The Handbook of Pragmatics*, 607–632, Blackwell.

Wood, S.N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, *1*(73), 3–36.

Yaneva, V., Ha, L.A., Evans, R., & Mitkov, R. (2018). Classifying referential and non-referential it using gaze. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 4896–4901, Brussels, Association for Computational Linguistics.

Zimmermann, M. (2008). Contrastive focus and emphasis. *Acta Linguistica Hungarica*, *55*(3–4), 347–360.

Zimmermann, M., & Onea, E. (2011). Focus marking and focus interpretation. *Lingua*, *121*(11), 1651–1670.